Proceedings of the 4th International Conference on Statistics: Theory and Applications (ICSTA'22)
Prague, Czech Republic – July 28- 30, 2022
Paper No. 162
DOI: 10.11159/icsta22.162

Clustering and Multidimensional Scaling for Individual Difference Extraction

Mika Sato-Ilic
Faculty of Engineering, Information and Systems, University of Tsukuba
Tennodai 1-1-1, Tsukuba, Japan
mika@risk.tsukuba.ac.jp

Abstract **-** This paper proposes methods to obtain difference among subjects by using the degree of reliability of each subject based on the results of fuzzy clustering and multidimensional scaling (MDS). In addition, new fuzzy clustering and MDS, including the weights of reliability scores, are proposed to classify subjects. When we observe data consisting of values of objects with respect to variables, and such data are observed over multiple subjects, capturing the difference among subjects is important in many fields. In this paper, the degree of reliability is obtained through the optimality of convex clustering. Based on this idea, it is shown that the same difference over the subjects can be obtained, regardless of the difference in obtained latent structures, which are the result of dynamic fuzzy clustering and the result of MDS by a numerical example. From this, we show the robustness of the proposed reliability concerning the variety of the obtained latent structures of data.

Keywords**:** fuzzy clustering, convex clustering, multidimensional scaling

1. Introduction

Systems based on the user's personal specifications are regarded as necessary in various fields such as long-term care, education, and medical care. Particularly in recent years, how these systems can be effectively used in human life is important for enriching the digital society of the future. In order to realize such a system, it is necessary to extract easily and quickly the difference between subjects from the obtained data and apply the difference efficiently to the results of analysis of the data to improve the result. For example, recently, we obtained many kinds of data of objects with respect to variables through several subjects. Several persons (subjects) wore sensor units on their bodies and did the same daily activity for a period of time. Then we can obtain data that consists of times (objects) in the period, measurement attributes (variables) of the sensors, and several subjects. When we implement a system to assist daily movement for a subject, it would be useful to learn the difference of the observed data among the subjects easily and quickly to improve the individually different results of analysis of the data. Such a system will be useful for adjusting the elderly user-friendly assistance system for daily activities considering the user's differences, such as a custom-made system.

Therefore, this paper proposes weighted individuality-based fuzzy clustering and MDS. In both of these, weights showing the degree of reliability for the obtained latent structures resulting from dynamic fuzzy clustering for each subject and the result of MDS for each subject are defined using the idea of the objective function of convex clustering. [1] Usually, the objective function [2], [3] is used to obtain a clustering result. However, in this study, conversely, the clustering result is given in the objective function as the latent structure of data, which is a result of dynamic fuzzy clustering [4] or a result of MDS [5], [6], [7] for each subject to measure how much each latent structure of each subject fits the optimality of convex clustering. Then, the score of the objective function can show the degree of reliability for the obtained latent structures of data at each subject, such as the result of the dynamic fuzzy clustering and the result of the MDS at each subject. Therefore, by including these scores as the degree of reliability of the original results, we can improve the results by considering the individual difference over the subjects.

The values of the reliability scores are mathematically comparable in the same criterion based on the objective function of convex clustering, so we can investigate the robustness of individual scores by using different methods, such as dynamic fuzzy clustering or MDS, in indicating differences among subjects. A numerical example shows a better performance of the robustness of the proposed degree of reliability for the difference of the latent structures obtained from data over the subjects.

This paper consists as follows: In section 2, we describe the dynamic fuzzy clustering and related ordinary fuzzy clustering methods. [4], [8], [9] In section 3, metric multidimensional scaling is described. [5], [6], [7] In section 4, the definition of the degree of reliability is described [1], and methods for obtaining the results of the fuzzy clustering and multidimensional scaling (MDS), which show the difference among the subjects, are proposed. In section 5, numerical examples using the proposed methods are explained, and in section 6, several conclusions are described.

2. Fuzzy Clustering

Suppose X be a given data matrix consisting of n objects and p variables as follows:

$$X = \begin{pmatrix} x_{11} & \cdots & x_{1p} \\ \vdots & \vdots & \vdots \\ x_{n1} & \cdots & x_{np} \end{pmatrix} = \begin{pmatrix} \boldsymbol{x}_1 \\ \vdots \\ \boldsymbol{x}_n \end{pmatrix}, \qquad \boldsymbol{x}_i = \left(x_{i1}, \cdots, x_{ip}\right), i = 1, \cdots, n, \tag{1}$$

where, $\boldsymbol{x}_i$ shows a vector for an object i. The purpose of clustering is to classify the n objects in (1) into K clusters. The state of fuzzy clustering is represented by a partition matrix:

$$U = (u_{ik}), \qquad i = 1, \cdots, n, \ \ k = 1, \cdots, K,$$

where u_{ik} shows the degree of belongingness of an object i to a cluster k. In general, u_{ik} satisfies the following conditions:

$$u_{ik} \in [0,1], \quad \sum_{k=1}^{K} u_{ik} = 1. \tag{2}$$

The fuzzy c-means (FCM) method [8] is one of the methods of fuzzy clustering. The purpose of this clustering method is to obtain solutions U and $\boldsymbol{v}_k = \left(v_{k1}, \cdots, v_{kp}\right), k = 1, \cdots, K$ which minimize the following weighted within-class sum of squares:

$$J(U, \boldsymbol{v}_1, \cdots, \boldsymbol{v}_K) = \sum_{i=1}^{n} \sum_{k=1}^{K} u_{ik}^{m} \, d^2(\boldsymbol{x}_i, \boldsymbol{v}_k), \tag{3}$$

where $\boldsymbol{v}_k = (v_{k1}, \cdots, v_{kp})$ denotes the values of the centroid of a cluster k, $\boldsymbol{x}_i = (x_{i1}, \cdots, x_{ip})$ is i-th object, and $d^2(\boldsymbol{x}_i, \boldsymbol{v}_k)$ is the square Euclidean distance between $\boldsymbol{x}_i$ and $\boldsymbol{v}_k$. The exponent m which determines the degree of fuzziness of the clustering is chosen from $(1, \infty)$ in advance. By minimizing (3), we obtain the solutions U and $\boldsymbol{v}_1, \cdots, \boldsymbol{v}_K$ as follows:

$$u_{ik} = \left[\sum_{j=1}^{K} \left(\frac{d(\boldsymbol{x}_i, \boldsymbol{v}_k)}{d\left(\boldsymbol{x}_i, \boldsymbol{v}_j\right)} \right)^{\frac{2}{m-1}} \right]^{-1}, \qquad \boldsymbol{v}_k = \frac{\sum_{i=1}^{n} u_{ik}^{m} \boldsymbol{x}_i}{\sum_{i=1}^{n} u_{ik}^{m}}, \qquad i = 1, \cdots, n, \qquad k = 1, \cdots, K.$$

Suppose Z_t be a given data matrix consisting of n objects with respect to p variables at a subject t called a 3-way data and shown as follows:

$$Z_t = \left(z_{ir}^{(t)}\right), \qquad i = 1, \cdots, n, \qquad r = 1, \cdots, p, \qquad t = 1, \cdots, T. \tag{4}$$

To obtain the same clusters over the T subjects, the following $nT \times p$ super matrix $\tilde{Z}$ is created.

$$\tilde{Z} = \begin{pmatrix} Z_1 \\ \vdots \\ Z_T \end{pmatrix} = \left(\tilde{z}_{jr}\right), \quad j = 1, \cdots, nT, \quad r = 1, \cdots, p. \tag{5}$$

The purpose of the dynamic fuzzy clustering [4] is to classify the nT objects into K clusters. The state of the fuzzy clustering is represented by a partition matrix:

$$\tilde{U} = \begin{pmatrix} U_1 \\ \vdots \\ U_T \end{pmatrix} = \left(\tilde{u}_{jk}\right), \qquad U_t = \left(u_{ik}^{(t)}\right), \qquad j = 1, \cdots, nT, \qquad i = 1, \cdots, n, \qquad k = 1, \cdots, K, \qquad t = 1, \cdots, T, \tag{6}$$

where $\tilde{u}_{jk}$ is a degree of belongingness of an object j which is shown as $\tilde{\boldsymbol{z}}_j = (\tilde{z}_{j1}, \cdots, \tilde{z}_{jp})$ to a fuzzy cluster k and $u_{ik}^{(t)}$ is a degree of belongingness of an object i to the same fuzzy cluster k at a subject t. From (6), the obtained K fuzzy clusters are the same over T subjects. In general, $\tilde{u}_{jk}$ satisfies the following conditions:

$$\tilde{u}_{jk} \in [0,1], \qquad \sum_{k=1}^{K} \tilde{u}_{jk} = 1, \qquad j = 1, \cdots, nT. \tag{7}$$

Then objective function of dynamic fuzzy clustering is defined by using FCM as follows:

$$J(\widetilde{U}, \boldsymbol{g}_1, \cdots, \boldsymbol{g}_K) = \sum_{j=1}^{nT} \sum_{k=1}^{K} \tilde{u}_{jk}^m \, d^2(\tilde{\boldsymbol{z}}_j, \boldsymbol{g}_k), \tag{8}$$

where $\boldsymbol{g}_k = (g_{k1}, \cdots, g_{kp})$ denotes the values of the center of a cluster k, $d^2(\tilde{\boldsymbol{z}}_j, \boldsymbol{g}_k)$ is the squared Euclidean distance between $\tilde{\boldsymbol{z}}_j$ and $\boldsymbol{g}_k$. The exponent m that determines the degree of fuzziness of the clustering is chosen from $(1, \infty)$ in advance. By minimizing the objective function in (8) under the conditions in (7), we obtain the solutions $\widetilde{U}, \boldsymbol{g}_1, \cdots, \boldsymbol{g}_K$.

If observed data is dissimilarity data, $\Delta = (\delta_{ij})$, where δ_{ij} shows dissimilarity between objects i and j, we use a fuzzy clustering method named FANNY algorithm. [9] The objective function of FANNY algorithm is defined as follows:

$$J(U) = \sum_{k=1}^{K} \left(\sum_{i=1}^{n} \sum_{j=1}^{n} u_{ik}^m u_{jk}^m \delta_{ij} \Big/ \left(2 \sum_{l=1}^{n} u_{lk}^m \right) \right). \tag{9}$$

By minimizing (9) under the conditions in (2), we obtain the solution U.

3. Multidimensional Scaling

From (1), when we assume that there are n objects denoted by n vectors in p dimensional space, the purpose of the Multidimensional scaling (MDS) is to obtain n points in a r ($r < p$) lower dimensional space with holding approximately the same similarity (or dissimilarity) relationship among objects in the p dimensional space. Then we can reduce the number of dimensions for capturing efficient information from observed data by representing the data structure in a lower dimensional spatial space. As a metric MDS, the following model has been proposed.

$$\delta_{ij} = \left\{ \sum_{\lambda=1}^{r} (\tilde{x}_{i\lambda} - \tilde{x}_{j\lambda})^2 \right\}^{\frac{1}{2}} + \varepsilon_{ij}, \quad i, j = 1, \cdots, n. \tag{10}$$

In (10), $\tilde{x}_{i\lambda}$ is a point of an object i with respect to dimension λ in r $(r < p)$ dimensional configuration space. ε_{ij} is an error. If we observe data of objects with respect to variables in (1), then the dissimilarity, δ_{ij}, is usually calculated by using Euclidean distance between objects i and j as follows:

$$\delta_{ij} \equiv \left\{ \sum_{a=1}^{p} \left(x_{ia} - x_{ja} \right)^2 \right\}^{\frac{1}{2}}, \quad i, j = 1, \cdots, n, \tag{11}$$

and apply the calculated dissimilarity in (11) to the MDS model in (10). That is, MDS finds r dimensional scaling (coordinate) $(\tilde{x}_{i1}, \cdots, \tilde{x}_{ir})$ and throws light on the structure of similarity relationship among the objects by representing the δ_{ij} in (10) as the distance between a point $(\tilde{x}_{i1}, \cdots, \tilde{x}_{ir})$ and a point$(\tilde{x}_{j1}, \cdots, \tilde{x}_{jr})$ in r dimensional space. Suppose $\varepsilon_{ij} = 0$, $\forall i, j$ in (10), then (10) can be rewritten as follows:

$$\Delta^2 = \mathbf{11}' diag(\tilde{X}\tilde{X}') - 2(\tilde{X}\tilde{X}') + diag(\tilde{X}\tilde{X}')\mathbf{11}', \tag{12}$$

where $\Delta^2$is a $n \times n$ matrix whose (i, j) element is δ_{ij}^2, and $\mathbf{1}$ is a vector whose n elements are all 1, and $\tilde{X}$ is a $n \times r$ matrix whose(i, λ) element is $\tilde{x}_{i\lambda}$, that is, $\Delta^2 = (\delta_{ij}^2)$, $\mathbf{1} = (1, \cdots, 1)'$, $\tilde{X} = (\tilde{x}_{i\lambda}), i, j = 1, \cdots, n,\ \lambda = 1, \cdots, r$, and $diag(A)$ means a diagonal matrix whose diagonal elements are consisted of diagonal elements of A. From the well-known Young-Householder transformation [10], which is a foundation of the MDS, Δ^2 in (12) can be transformed as follows:

$$P = -\frac{1}{2} J \Delta^2 J = \tilde{X}\tilde{X}', \tag{13}$$

where $P = (p_{ij}),\ i, j = 1, \cdots, n$. Matrix J is a symmetric matrix whose diagonal elements are $1 - 1/n$ and non-diagonal elements are $-1/n$ which means centering operation for each column of $\tilde{X}$, that is the following condition

$$\sum_{i=1}^{n} \tilde{x}_{i\lambda} = 0, \qquad \forall \lambda$$

is satisfied to fix an origin as 0 for all n dimensions in the obtained coordinate space. By using eigenvalue decomposition of P in (13), P can be represented as follows:

$$P = H\Lambda H' = H\Lambda^{\frac{1}{2}}\Lambda^{\frac{1}{2}}H' = \left(H\Lambda^{\frac{1}{2}}\right)\left(H\Lambda^{\frac{1}{2}}\right)', \tag{14}$$

where $H = \left(h_{ij}\right) = (\boldsymbol{h}_1, \cdots, \boldsymbol{h}_n)$, $\boldsymbol{h}_\lambda = (h_{1\lambda}, \cdots, h_{n\lambda})'$, $i, j = 1, \cdots, n$, and Λ is a diagonal matrix whose diagonal elements are eigen values $\lambda_1, \cdots, \lambda_n$ and satisfy $\lambda_1 > \cdots > \lambda_n$. $\Lambda^{\frac{1}{2}}$ is a diagonal matrix whose diagonal elements are eigen values $\sqrt{\lambda_1}, \cdots, \sqrt{\lambda_n}$. H is a matrix whose column vectors are eigen vectors $\boldsymbol{h}_1, \cdots, \boldsymbol{h}_n$ corresponding to eigen values $\lambda_1, \cdots, \lambda_n$. When $r < p < n$ and eigen values $\lambda_{r+1}, \cdots, \lambda_n$ are close to 0, that is, dimensions $r+1$ to n do not have explanatory power for the given dissimilarity data, (14) can be approximately represented as follows by using fewer r dimensions:

$$P = H\Lambda H' \approx \widetilde{H}\widetilde{\Lambda}\widetilde{H}' = \widetilde{H}\widetilde{\Lambda}^{\frac{1}{2}}\widetilde{\Lambda}^{\frac{1}{2}}\widetilde{H}' = \left(\widetilde{H}\widetilde{\Lambda}^{\frac{1}{2}}\right)\left(\widetilde{H}\widetilde{\Lambda}^{\frac{1}{2}}\right)', \tag{15}$$

where $\widetilde{H} = (h_{i\lambda}) = (\boldsymbol{h}_1, \cdots, \boldsymbol{h}_r)$, $i = 1, \cdots, n$, $\lambda = 1, \cdots, r$, and $\widetilde{\Lambda}$ is a diagonal matrix whose diagonal elements are eigen values $\delta_1, \cdots, \delta_r$. $\widetilde{\Lambda}^{\frac{1}{2}}$ is a diagonal matrix whose diagonal elements are eigen values $\sqrt{\delta_1}, \cdots, \sqrt{\delta_r}$. From (13) and (15), $\tilde{X}$ can be estimated as follows:

$$\hat{X} = \widetilde{H}\widetilde{\Lambda}^{\frac{1}{2}}, \tag{16}$$

where $\hat{X} = (\hat{x}_{i\lambda}), i = 1, \cdots, n$, $\lambda = 1, \cdots, r$, and the elements of $\hat{X}$ show values of coordinate of n objects in r dimensional space when $r < p < n$. Therefore, the elements of $\hat{X}$ in (16) are the estimate of $\tilde{x}_{i\lambda}$ in the model of MDS in (10).

4. Weighted Individuality-Based Fuzzy Clustering and Multidimensional Scaling

Convex clustering is a type of clustering method in which we obtain clustering results by solving a convex optimization problem. The idea is based on the sparsity regularization of regression. [11]

Suppose $\boldsymbol{q}_i$ be a centroid for the cluster containing $\boldsymbol{x}_i$ shown in (1) as follows:

$$Q = \begin{pmatrix} q_{11} & \cdots & q_{1p} \\ \vdots & \vdots & \vdots \\ q_{n1} & \cdots & q_{np} \end{pmatrix} = \begin{pmatrix} \boldsymbol{q}_1 \\ \vdots \\ \boldsymbol{q}_n \end{pmatrix}, \qquad \boldsymbol{q}_i = \left(q_{i1}, \cdots, q_{ip}\right), \qquad i = 1, \cdots, n.$$

The purpose of convex clustering is to obtain $\boldsymbol{q}_i$ which minimizes the following function:

$$F_\omega(Q) = \frac{1}{2}\sum_{i=1}^{n} d^2(\boldsymbol{x}_i, \boldsymbol{q}_i) + \omega \sum_{i<j} d^2(\boldsymbol{q}_i, \boldsymbol{q}_j). \tag{17}$$

In (17), $d^2(\boldsymbol{x}_i, \boldsymbol{q}_i)$ is the square Euclidean distance between $\boldsymbol{x}_i$ and $\boldsymbol{q}_i$ and $d^2(\boldsymbol{q}_i, \boldsymbol{q}_j)$ is the square Euclidean distance between $\boldsymbol{q}_i$ and $\boldsymbol{q}_j$. ω is a given parameter in which if $\omega = 0$, then we simply obtain a solution as $\boldsymbol{x}_i = \boldsymbol{q}_i, \forall i$, that is, we obtain n clusters in which only one object belongs to one cluster. If $\omega \to \infty$, then we obtain one cluster containing all objects. When we obtain the optimum solution, several values of the square Euclidean distances between a pair of centroids are zeros. For example, if $d^2(\boldsymbol{q}_s, \boldsymbol{q}_l) = 0$, then $\boldsymbol{x}_s$ and $\boldsymbol{x}_l$ belong to the same cluster. This shows a merging process of the clustering.

Suppose we observe 3-way data in (4). Then based on the objective function of convex clustering in (17) when $\omega = 1$, the weight of subject t based on a result of dynamic FCM is defined as follows [1]:

$$w_{FCM}^{(t)} = \frac{1}{2}\sum_{i=1}^{n} d^2\left(\boldsymbol{z}_i^{(t)}, \boldsymbol{u}_i^{(t)}\right) + \sum_{i<j} d^2\left(\boldsymbol{u}_i^{(t)}, \boldsymbol{u}_j^{(t)}\right), \quad t = 1, \cdots, T, \tag{18}$$

where $\boldsymbol{z}_i^{(t)} = (z_{i1}^{(t)}, \cdots, z_{ip}^{(t)})$ in (4) is a data of an object i at t-th subject and $\boldsymbol{u}_i^{(t)} = (u_{i1}^{(t)}, \cdots, u_{ip}^{(t)})$ is a result of dynamic FCM for an object i at t-th subject when $K = p$ in (6), when we apply data $\tilde{Z}$ in (5) to the dynamic fuzzy clustering method in (8).

Next, we define the weight of subject t based on a result of MDS as follows [1]:

$$\widetilde{w}_{MDS}^{(t)} = \frac{1}{2}\sum_{i=1}^{n} d^2\left(\boldsymbol{z}_i^{(t)}, \hat{\boldsymbol{x}}_i^{(t)}\right) + \sum_{i<j} d^2\left(\hat{\boldsymbol{x}}_i^{(t)}, \hat{\boldsymbol{x}}_j^{(t)}\right), \qquad t = 1, \cdots, T, \tag{19}$$

where and $\hat{\boldsymbol{x}}_i^{(t)} = (\hat{x}_{i1}^{(t)}, \cdots, \hat{x}_{ip}^{(t)})$ is a result of MDS for an object i at t-th subject in (16) when $r = p$ in the case when we calculate Euclidean distance between objects for each subject t by using Z_t in (4), and apply to the MDS in (10).

In (18) and (19), $U_t = \begin{pmatrix} \boldsymbol{u}_1^{(t)} \\ \vdots \\ \boldsymbol{u}_n^{(t)} \end{pmatrix}$ and $\hat{X}^{(t)} = \begin{pmatrix} \hat{\boldsymbol{x}}_1^{(t)} \\ \vdots \\ \hat{\boldsymbol{x}}_n^{(t)} \end{pmatrix}$ are representative of t-th subject obtained from different latent structures of the data, respectively, and the values of $w_{FCM}^{(t)}$ and $\widetilde{w}_{MDS}^{(t)}$ show how much the representative of t-th subject fits the original each subject's data which is shown as Z_t in (4), and how much the representative shows well-classified status as the optimality of the convex clustering. That is, smaller values of $w_{FCM}^{(t)}$ and $\widetilde{w}_{MDS}^{(t)}$ mean that the obtained latent structures as the results of the dynamic fuzzy clustering and MDS for the subject t has higher reliability to the real data at the subject t.

Moreover, we can compare values for each pair $(w_{FCM}^{(t)}, \widetilde{w}_{MDS}^{(t)}), t = 1, \cdots, T$. From this, we can investigate the robustness of the weights in indicating differences among subjects, by using different representatives, which are different latent structures, U_t and $\hat{X}^{(t)}$ from the same subject t. Then the difference between Z_t and $Z_{t'}$ which are data for subjects t and t' in (4) is calculated as follows:

$$\hat{\delta}_{tt'} = \sum_{i=1}^{n} \sum_{r=1}^{p} \left(z_{ir}^{(t)} - z_{ir}^{(t')} \right)^2, \quad t, t' = 1, \cdots, T, \quad t \neq t'. \tag{20}$$

We normalized the dissimilarity in (20) and obtain the normalized dissimilarity as $\overline{\Delta} = (\bar{\delta}_{tt'})$. To avoid negative values in $\overline{\Delta} = (\bar{\delta}_{tt'})$, we recalculate as follows:

$$\widetilde{\Delta} = (\tilde{\delta}_{tt'}), \qquad \tilde{\delta}_{tt'} = \bar{\delta}_{tt'} + abs\left(\min_{t,t'} \bar{\delta}_{tt'} \right), \qquad t, t' = 1, \cdots, T, \quad t \neq t', \tag{21}$$

where $abs(*)$ means absolute value of $*$. By applying $\widetilde{\Delta}$ in (21) to FANNY in (9), we obtain the result of the FANNY as follows:

$$\widehat{U} = (\hat{u}_{tk}), t = 1, \cdots, T, k = 1, \cdots, K, \tag{22}$$

under the conditions as follows:

$$\hat{u}_{tk} \in [0,1], \quad \sum_{k=1}^{K} \hat{u}_{tk} = 1, \tag{23}$$

where $\hat{u}_{tk}$ shows a degree of belongingness of a subject t to a cluster k. However, the result in (22) does not consider the degree of reliability for each subject to match the individually different data structure based on the weight in (18). In addition, from the conditions in (23), the result in (22) has constraints that reduce the result's explanatory power. To overcome this problem, we propose the following weighted individuality based fuzzy clustering result as follows:

$$W\widehat{U} = (\bar{u}_{tk}), W = \begin{pmatrix} w^{(1)} & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & w^{(T)} \end{pmatrix}, \bar{u}_{tk} = \quad w^{(t)} \hat{u}_{tk}, w^{(t)} = \frac{w_{FCM}^{(t)}}{\sum_{t=1}^{T} w_{FCM}^{(t)}}, \qquad t = 1, \cdots, T, k = 1, \cdots, K, \tag{24}$$

where W is a diagonal matrix whose diagonal elements are $w^{(1)}, \cdots, w^{(T)}$. Note that $\bar{u}_{tk}$ in (24) does not have the constraints in (23). Also, from (24), it can be seen that $\bar{u}_{tk}$ considers the normalized degree of reliability for subject t which is represented by $w^{(t)}$.

Next, we apply $\widetilde{\Delta}$ in (21) to MDS in (10) and obtain the result of MDS as follows:

$$\hat{\hat{X}} = (\hat{\hat{x}}_{t\lambda}), \qquad t = 1, \cdots, T, \qquad \lambda = 1, \cdots, r.$$

By considering the degree of reliability for each subject based on the MDS in (19), we define the following result for weighted individuality based MDS as follows:

$$\widetilde{W}\hat{\hat{X}} = (\bar{x}_{t\lambda}), \widetilde{W} = \begin{pmatrix} \widetilde{w}^{(1)} & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \widetilde{w}^{(T)} \end{pmatrix}, \bar{x}_{t\lambda} = \quad \widetilde{w}^{(t)} \hat{\hat{x}}_{t\lambda}, \widetilde{w}^{(t)} = \frac{\widetilde{w}_{MDS}^{(t)}}{\sum_{t=1}^{T} \widetilde{w}_{MDS}^{(t)}}, \qquad t = 1, \cdots, T, \lambda = 1, \cdots, r, \tag{25}$$

where $\widetilde{W}$ is a diagonal matrix whose diagonal elements are $\widetilde{w}^{(1)}, \cdots, \widetilde{w}^{(T)}$. From (25), it can be seen that $\bar{x}_{t\lambda}$ considers the normalized degree of reliability for subject t which is represented by $\widetilde{w}^{(t)}$.

5. Numerical Examples

We use a dataset of sensor data of daily and sports activities performed by 8 subjects with respect to 45 variables. [12], [13] The dataset consists of 19 regular daily and sports activities, and we select the activity of running on a treadmill at a speed of 8 km/h. For this activity, we use data which consisted of 7500 times, 45 variables. For the variables, the data was observed for 45 (5 body positions × 9 kinds of sensors) variables which consist of 5 body-worn sensor units positioned on the torso (T), right arm (RA), left arm (LA), right leg (RL), left leg (LL), and each sensor unit has 9 kinds of information including x-axial accelerometers (xacc), y-axial accelerometers (yacc), z-axial accelerometers (zacc), x-axial gyroscopes (xgyro), y-axial gyroscopes (ygyro), z-axial gyroscopes (zgyro), x-axial magnetometers (xmag), y-axial magnetometers (ymag), z-axial magnetometers (zmag).

Figure 1 shows a result of values of the weights which shows the degree of reliability for each subject. In this figure, the ordinate shows the values of $\{w^{(1)},\cdots,w^{(8)}\}$ in (24) and the abscissa shows the values of $\{\tilde{w}^{(1)},\cdots,\tilde{w}^{(8)}\}$ in (25). The point "pt" shows a point of t-th subject at the coordinate of $(\tilde{w}^{(t)}, w^{(t)}), t = 1,\cdots,8$. From this figure, the two kinds of weights based on different latent structures which are the result of MDS at t-th subject, $\hat{X}^{(t)}$, and the result of dynamic FCM at the same subject, U_t, are almost linear relationship in indicating difference among 8 subjects. In fact, the correlation between and over 8 subjects in this figure is 0.99. From this, the proposed weights for individual difference extraction based on the objective function of convex clustering shows robustness for the difference of the kinds of representative, which are latent structures of data obtained from different kinds of methods such as dynamic FCM and MDS.

Figure 2 shows the relationship between the difference of a pair of subjects of data in (20) and difference of the same pair of the subjects of the results of MDS. The difference of the same pair of the subjects t and t' of the results of MDS in (16) is calculated as follows:

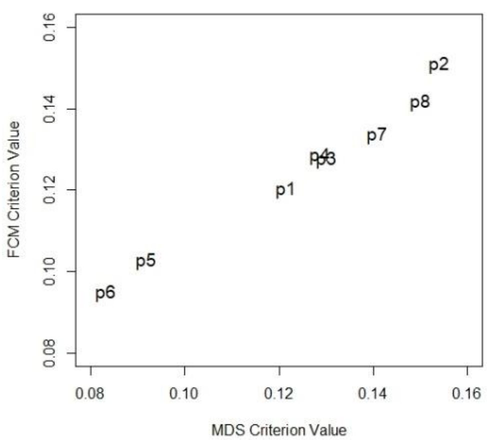
$$\check{\delta}_{tt'} = \sum_{i=1}^{7500}\sum_{\lambda=1}^{45}\left(\hat{x}_{i\lambda}^{(t)} - \hat{x}_{i\lambda}^{(t')}\right)^2, \quad t,t' = 1,\cdots,8, \quad t \neq t', \tag{26}$$

where $\hat{x}_{i\lambda}^{(t)}$ shows value of coordinate of an object i in λ dimension at t-th subject and $\hat{x}_{i\lambda}^{(t')}$ shows value of coordinate of an object i in λ dimension at t'-th subject. The point " $\mathrm{d}tt'$" is a point that shows a difference between subjects t and t' at the coordinate of $(\hat{\delta}_{tt'}, \check{\delta}_{tt'}), t,t' = 1,\cdots,8\ (t \neq t')$. From this figure, there is an almost monotone relationship between the two differences. This shows the validity of the result of MDS and the use of this result for the proposed reliability for indicating individuality in (19) and the proposed weighted individuality based MDS in (25).

Figure 3 shows the proposed result for weighted individuality based MDS in (25). From this figure, we can see that dimension 2 shows an ability to distinguish between male and female subjects. Subjects 1, 2, 6, and 7 are female subjects, and other subjects are male in this data. So, female subjects have the tendency of lower scores with respect to dimension 2, otherwise male subjects have an opposite tendency which has higher values with respect to dimension 2.

Figure 4 shows the result of FANNY in (22). In this figure, the abscissa shows the values of degree of belongingness of objects to cluster 1, and the ordinate is the values of degree of belongingness of objects to cluster 2. The number of clusters is assumed to be 2.

Figure 5 shows a result of the proposed weighted individuality-based FANNY applied dissimilarity among subjects of data in (24). From the comparison between figures 4 and 5, we can see two groups; one is a group of subjects 7 and 8, and another is a group of other subjects in both figures 4 and 5. However, from the constraints in (23), in figure 4, all the differences of 8 subjects are on the same line, and we cannot see the detailed difference over 8 subjects. And in figure 5, we can see the more detailed difference among 8 subjects compared with the original result of FANNY shown in figure 4. In particular, we can see the detailed difference among subjects 1 to 6 in figure 5 compared with the case of figure 4. From this comparison, we can obtain more detailed differences considering individual scores for each subject in the proposed weighted individuality-based fuzzy clustering.



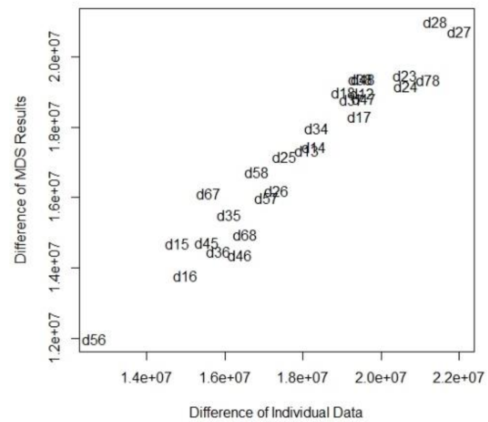
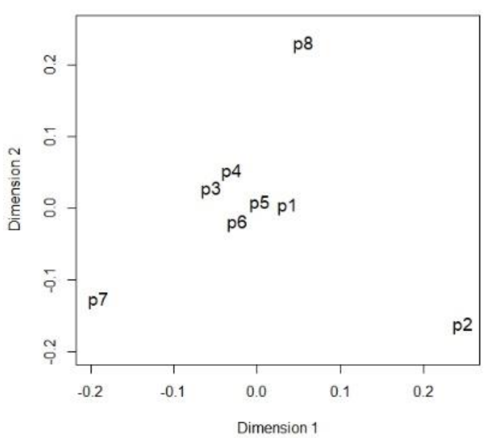


Fig. 1. Relationship between the two kinds of weights of subjects

Fig. 2. Relationship between the difference of data and the difference of MDS results for each pair of subjects



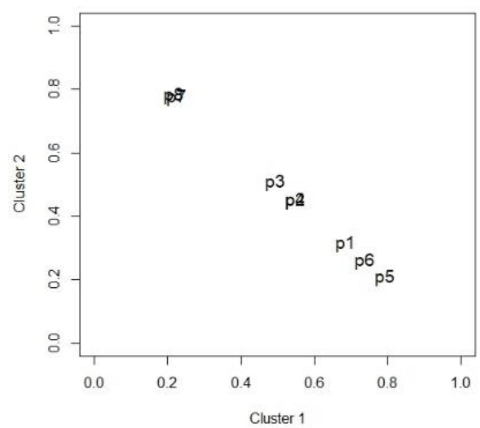


Fig. 3. Result of proposed weighted MDS applied dissimilarity among subjects

Fig. 4. Result of FANNY applied dissimilarity among subjects

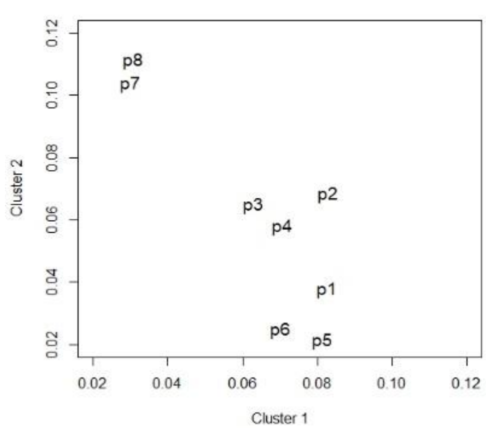


Fig. 5. Result of proposed weighted FANNY applied dissimilarity among subjects

6. Conclusions

This paper proposes methods to obtain results of fuzzy clustering and MDS for classifying subjects over the data consisted of objects, variables, and subjects. In these methods, we utilize the degree of reliability obtained by using the results of dynamic fuzzy clustering and MDS at each subject. This degree of reliability is defined as a degree that shows how much the obtained result of the fuzzy clustering or MDS fits the original subject's data as a latent structure of each subject's data and how much the obtained result shows well-classified status as the optimality of the convex clustering. Therefore, the validity of the degree of reliability of the result of dynamic fuzzy clustering and MDS at each subject is theoretically guaranteed.

In addition, we show the robustness of the degree of reliability with respect to the variety of the obtained latent structure of data by using a numerical example. We show that the degree of reliability used in a latent structure as a result of the dynamic fuzzy clustering is highly correlated with the degree of reliability used in a latent structure as a result of MDS. Other numerical examples show a better performance of the proposed methods compared with ordinary results. In particular, in the case of fuzzy clustering, the ordinary result cannot avoid the constraint in which the sum of the degree of belongingness of objects to clusters is 1 for a fixed object. Therefore, the explainable ability of the original data structure is also limited. By including the degree of reliability for each subject to the original fuzzy clustering result, we can avoid this constraint and obtain a more detailed result of the clustering of subjects.

The target data is many kinds of data of objects through several subjects; however, it is not limited to subjects. For example, if data will be obtained through several times, it would be useful to know the difference over the times, so we can rapidly detect the time of the fault. Therefore, in future studies, examinations of the proposed methods for various kinds of data are necessary to know the generalization performance. In addition, other kinds of latent structures obtained as representative for each subject should be applied to the proposed method to strengthen the investigation of the feature of the robustness of the proposed degree of reliability for the latent structure of each subject's data.

References

[1] M. Sato-Ilic, "Indicator for individuality of subjects based on similarity of objects," *Procedia Computer Science*, vol. 185, pp. 193–202, 2021.

[2] F. Lindsten, H. Ohlsson, and L. Ljung, "Just relax and come clustering! A convexification of k-means clustering," *Tech. rep.*, Linköpings universitet, 2011.

[3] E. C. Chi and K. Lange, "Splitting methods for convex clustering," *Journal of Computational and Graphical Statistics*, vol. 24, pp. 994-1013, 2015.

[4] M. Sato-Ilic, "Individual compositional cluster analysis," *Procedia Computer Science*, vol. 95, pp. 254-263, 2016.

[5] J. C. Gower, "Some distance properties of latent roots and vector methods used in multivariate analysis," *Biometrika*, vol. 53, pp. 325-338, 1966.

[6] J. B. Kruskal and M. Wish, *Multidimensional Scaling*, Sage Publications, 1978.

[7] W. S. Torgerson, *Theory and Methods of Scaling*, Wiley, 1958.

[8] J. C. Bezdek, *Pattern Recognition with Fuzzy Objective Function Algorithms*, Plenum, 1981.

[9] L. Kaufman and P.J. Rousseeuw, *Finding Groups in Data: An Introduction To Cluster Analysis*, Wiley, 2005.

[10] G. Young and A. S. Householder, "Discussion of a set of points in terms of their mutual distances," *Psychometrika*, vol. 3, pp. 19-22, 1938.

[11] M. Yuan and Y. Lin, "Model selection and estimation in regression with grouped variables," *Journal of the Royal Statistical Society*, Series B, vol. 68, pp. 49–67, 2006.

[12] K. Altun and B. Barshan, "Human activity recognition using inertial/magnetic sensor units", in A. A. Salah, T. Gevers, N. Sebe, A. Vinciarelli (eds) HBU 2010. LNCS 6219, Springer; Berlin, Heidelberg, pp. 38-51, 2010.

[13] UCI Machine Learning Repository, http://archive.ics.uci.edu/ml/index.html