https://doi.org/10.20965/jaciii.2018.p0054

Paper:

Two-Stage Clustering Based on Cluster Validity Measures

Yukihiro Hamasuna*, **Ryo Ozaki****, **and Yasunori Endo*****

*Department of Informatics, School of Science and Engineering, Kindai University,
3-4-1 Kowakae, Higashiosaka, Osaka 577-8502, Japan
E-mail: yhama@info.kindai.ac.jp
**Graduate School of Science and Engineering, Kindai University
3-4-1 Kowakae, Higashiosaka, Osaka 577-8502, Japan
E-mail: 1210370107b@gmail.com
***Faculty of Engineering, Information and Systems, University of Tsukuba,
1-1-1 Tennodai, Tsukuba, Ibaraki 305-8573, Japan
E-mail: endo@risk.tsukuba.ac.jp

[Received November 25, 2016; accepted October 7, 2017]

To handle a large-scale object, a two-stage clustering method has been previously proposed. The method generates a large number of clusters during the first stage and merges clusters during the second stage. In this paper, a novel two-stage clustering method is proposed by introducing cluster validity measures as the merging criterion during the second stage. The significant cluster validity measures used to evaluate cluster partitions and determine the suitable number of clusters act as the criteria for merging clusters. The performance of the proposed method based on six typical indices is compared with eight artificial datasets. These experiments show that a trace of the fuzzy covariance matrix $\boldsymbol{W}_{\mathrm{tr}}$ and its kernelization $\boldsymbol{KW}_{\mathrm{tr}}$ are quite effective when applying the proposed method, and obtain better results than the other indices.

Keywords: two-stage clustering, cluster validity measures, kernel method, *c*-means clustering, agglomerative hierarchical clustering

1. Introduction

Data clustering methods such as *k*-means [1] divide a set of objects into groups called clusters. Objects classified in the same cluster are considered similar, whereas those in different clusters are considered dissimilar. Hard *c*-means (HCM), variants of fuzzy *c*-means (FCM) [2, 3], and agglomerative hierarchical clustering (AHC) [4] are representative clustering methods. Data clustering methods have received significant attention in recent years for handling large-scale objects and extracting values from them.

In the field of data mining, a recent research topic is the handling of large-scale data obtained from real-world problems [5, 6]. A two-stage clustering method has also been proposed to tackle these problems [7–9]. In the first stage of the two-stage clustering method, a fast algorithm such as *k*-means generates a number of clusters. Then, in the second stage, clusters obtained in the first stage are clustered using an AHC procedure or other clustering methods. Using two-stage clustering, time-consuming clustering methods such as AHC, kernel clustering, and spectral clustering can be applied to large-scale data.

The strong point of two-stage clustering is the strategy applied in the second stage. In previous studies, significant clustering methods have been used in the second stage [7–9]. We introduced cluster validity measures [10–13] in the second stage as a novel approach to two-stage clustering and proposed a framework of two-stage clustering based on cluster validity measures [14]. Cluster validity measures have also been actively studied for the purpose of evaluating cluster partitions and determining the suitable number of clusters [10–13, 15]. For this reason, we have considered these cluster validity measures as possible criterioa to merge clusters in the second stage [14]. In previous studies, numerical experiments indicated the effectiveness of two-stage clustering based on a trace of the fuzzy covariance matrix (W_{tr}) [11], Xie-Beni's index (XB), and partition coefficients [13]. However, experiments also showed that conventional indices including W_{tr} and XB are not suitable merging criteria for significant datasets consisting of different cluster shapes or sizes. We introduced Fukuyama-Sugeno's index (FS) [16] and several kernelized measures [15] to handle various types of datasets that consist of clusters of several shapes or sizes. We also compared these indices to others used in the previous study through several numerical experiments to show the effectiveness of the two-stage clustering based on cluster validity measures.

The reminder of this paper is organized as follows: In Section 2, we introduce the notation, *c*-means clustering, agglomerative hierarchical clustering, and cluster validity measures. In Section 3, we describe a method of two-stage clustering based on cluster validity measures. In Section 4, we conduct experiments to show the effectiveness of the proposed method. In Section 5, we provide some concluding remarks regarding this research.

2. Preliminaries

A set of objects to be clustered is given and denoted by $X = \{x_1, \ldots, x_n\}$, in which $x_k\,(k = 1, \ldots, n)$ is an object. In most cases, each object x_k is a vector in p-dimensional Euclidean space $\Re^p$, that is, an object $x_k \in \Re^p$. A cluster is denoted by G_i, and a collection of clusters is given by $\mathscr{G} = \{G_1, \ldots, G_c\}$. A cluster center of G_i is denoted by $v_i \in \Re^p$, and a set of v_i is given by $V = \{v_1, \ldots, v_c\}$. The membership degree of x_k belonging to G_i and a partition matrix is denoted as u_{ki}, and $U = (u_{ki})_{1 \le k \le n,\ 1 \le i \le c}$.

2.1. *c*-Means Clustering

k-means is a well-known clustering method, and is also called hard *c*-means (HCM) [1], in contrast with FCM [2, 3]. These algorithms divide a set of objects into clusters by optimizing an objective function under the constraint on the membership degree.

The objective function of HCM, denoted as J_h, is as follows:

$$J_h(U,V) = \sum_{k=1}^{n}\sum_{i=1}^{c} u_{ki} d_{ki},$$

where d_{ki} is the dissimilarity between an object x_k and cluster center v_i. The squared L_2-norm $d_{ki} = \|x_k - v_i\|^2$ is a typical dissimilarity in HCM and FCM.

The constraint on membership degree u_{ki} is as follows:

$$\mathscr{U}_h = \left\{ (u_{ki}) : u_{ki} \in \{0,1\},\ \sum_{i=1}^{c} u_{ki} = 1,\ {}^{\forall} k \right\}.$$

The optimal solutions for u_{ki} and v_i of HCM are as follows:

$$u_{ki} = \begin{cases} 1 & (\ i = \arg\min_l \|x_k - v_l\|^2\ ) \\ 0 & (\ \text{otherwise}\ ) \end{cases}, \quad \ldots \quad (1)$$

$$v_i = \frac{\sum_{k=1}^{n} u_{ki} x_k}{\sum_{k=1}^{n} u_{ki}}. \quad \ldots \quad (2)$$

FCM is also based on optimizing an objective function under the constraint for the membership degree.

In the following two objective functions, J_s is representative of FCM.

$$J_s(U,V) = \sum_{k=1}^{n}\sum_{i=1}^{c} (u_{ki})^m d_{ki}.$$

Here, $m > 1$ is the fuzzification parameter.

In addition to J_s [2], entropy-based fuzzy *c*-means clustering is also representative of FCM [3].

The constraint on membership degree u_{ki} for FCM is as follows:

$$\mathscr{U}_f = \left\{ (u_{ki}) : u_{ki} \in [0,1],\ \sum_{i=1}^{c} u_{ki} = 1,\ {}^{\forall} k \right\}.$$

The optimal solutions for u_{ki} and v_i derived from J_s are as follows:

$$u_{ki} = \frac{\left(\dfrac{1}{d_{ki}}\right)^{\frac{1}{m-1}}}{\sum_{l=1}^{c}\left(\dfrac{1}{d_{kl}}\right)^{\frac{1}{m-1}}}, \quad \ldots \quad (3)$$

$$v_i = \frac{\sum_{k=1}^{n} (u_{ki})^m x_k}{\sum_{k=1}^{n} (u_{ki})^m}. \quad \ldots \quad (4)$$

Algorithm 1 HCM and FCM

STEP 1 Set initial cluster centers and parameters.

STEP 2 Calculate $u_{ki} \in U$ using Eq. **A**.

STEP 3 Calculate $v_i \in V$ using Eq. **B**.

STEP 4 If the convergence criterion is satisfied, stop. Otherwise go back to **STEP 2**.

Table 1. Optimal solutions of HCM and FCM.

Algorithm	Eq. **A**	Eq. **B**
HCM	(1)	(2)
FCM	(3)	(4)

The algorithm of HCM and FCM is summarized as **Algorithm 1**.

The number of repetitions, the convergence of each variable, or the convergence of an objective function are used as the convergence criterion in **STEP 4**. The optimal solutions Eqs. **A** and **B** used in each algorithm are shown in **Table 1**.

2.2. Agglomerative Hierarchical Clustering

Along with *c*-means clustering, agglomerative hierarchical clustering (AHC) is also a representative clustering algorithm as well as *c*-means clustering [4]. The AHC algorithm merges clusters sequentially and outputs a dendrogram. Crisp clusters obtained by AHC are disjointed and their union is a set of objects as follows:

$$G_i \cap G_j = \emptyset \quad (i \neq j),$$

$$\bigcup_{i=1}^{c} G_i = X.$$

A dissimilarity $d(G, G')$ $(G, G' \in \mathscr{G})$ is defined as the criterion that measures the nearness of two clusters. We describe a general procedure of AHC as **Algorithm 2**.

There are several methods in the AHC procedure for updating the dissimilarity, including single linkage, complete linkage, average linkage, centroid method, and Ward's method. The fundamental characteristic of these methods is omitted here; see [4] for details of these methods.

Algorithm 2 AHC

AHC 1 Assume that the initial clusters are given by

$$\mathscr{G} = \{G_1, G_2, \ldots, G_n\}$$

Set $C := n$. (C is the number of clusters and n is the initial number of clusters)

Calculate $d(G, G')$ for all pairs $G, G' \in \mathscr{G}$.

AHC2 Search the pair of minimum dissimilarities:

$$(G_p, G_q) = \arg \min_{G, G' \in \mathscr{G}} d(G, G').$$

Merge: $G_r := G_p \cup G_q$.

Add G_r to $\mathscr{G}$ and delete G_p and G_q from $\mathscr{G}$.

$C := C - 1$.

If $C = 1$ then stop and output the dendrogram.

AHC 3 Update $d(G_r, G'')$ for all other $G'' \in \mathscr{G}$.

Go to **AHC 2**.

2.3. Cluster Validity Measures

Cluster validity measures are used to evaluate clustering partitions and determine the number of clusters [11–13, 15, 16]. To date, many cluster validity measures and their extensions have been proposed and studied [13, 15]. Two types of cluster validity measures are considered herein, namely, validity indices involving only the membership degree, and validity indices involving geometric features.

2.3.1. Partition Coefficients

First, we introduce the following three indices involving only the membership degree, that is, partition coefficient PC, partition entropy PE, and modification of partition coefficient MPC [13].

$$PC = \frac{1}{n} \sum_{k=1}^{n} \sum_{i=1}^{c} (u_{ki})^2,$$

$$PE = -\frac{1}{n} \sum_{k=1}^{n} \sum_{i=1}^{c} u_{ki} \log u_{ki},$$

$$MPC = 1 - \frac{c}{c-1} (1 - PC).$$

If the value of PC or MPC is large, the cluster partition is considered to be good, whereas if the value of PE is small, the result is also considered to be good.

2.3.2. Fuzzy Covariance Matrix Based Indices

Gath-Geva's index [11] is based on fuzzy covariance matrix F_i. Two types of indices are considered, that is, the sum of the determinants of F_i and the sum of the traces of F_i, which are as follows:

$$W_{\mathrm{det}} = \sum_{i=1}^{c} \sqrt{\det F_i},$$

$$W_{\mathrm{tr}} = \sum_{i=1}^{c} \mathrm{tr} F_i.$$

The fuzzy covariance matrix F_i is defined as follows:

$$F_i = \frac{\sum_{k=1}^{n} (u_{ki})^m (x_k - v_i)(x_k - v_i)^T}{\sum_{k=1}^{n} (u_{ki})^m}. \quad \ldots \ldots \quad (5)$$

Here, m is a fuzzified parameter used in FCM [2]. For HCM, fuzzified parameter $m = 1$ in F_i. If the value of W_{det} or W_{tr} is small, the clustering partition is considered to be good.

2.3.3. Xie-Beni's Index

Xie-Beni's index (XB) [12] is constructed by considering the objective function and the minimum dissimilarity between cluster centers, and can be described as follows:

$$XB = \frac{\sum_{k=1}^{n} \sum_{i=1}^{c} (u_{ki})^m \|x_k - v_i\|^2}{n \min_{1 \le i, j, \le c,\ i \neq j} \|v_i - v_j\|^2}. \quad \ldots \ldots \quad (6)$$

Here, m is also a fuzzified parameter. For HCM, $m = 1$ in the numerator as well as for F_i. If the value of XB is small, the clustering result is considered to be good.

2.3.4. Fukuyama-Sugeno's Index

Fukuyama-Sugeno's index (FS) is constructed by considering the objective function and the dissimilarity of each cluster centers and the center of the dataset and can be described as follows:

$$FS = \sum_{k=1}^{n} \sum_{i=1}^{c} (u_{ki})^m \left[\|x_k - v_i\|^2 - \|v_i - \tilde{v}\|^2 \right] \quad \ldots \quad (7)$$

where $\tilde{v}$ is the center of the dataset.

$$\tilde{v} = \frac{1}{n} \sum_{k=1}^{n} x_k.$$

If the value of FS is small, the clustering result is considered to be good.

2.3.5. Kernelized Indices

Clustering methods with a kernel function generate a cluster partition with a non-linear boundary [17]. The kernelized cluster validity measures are also studied to evaluate cluster partitions with a non-linear boundary and the nearness of each object in a high-dimensional feature space [15].

First, we define symbols to introduce kernel functions. Here, $\phi : \Re^p \to \Re^s (p \ll s)$ indicates mapping from the input space $\Re^p$ to high dimensional feature space $\Re^s$. An object in the feature space is denoted as $\phi(x_k) \in \Re^s$. Kernel function $K : \Re^p \times \Re^p \to \Re$ satisfies the following relation:

$$K(x,y) = \langle \phi(x), \phi(y) \rangle.$$

Note that ϕ is not explicit. The inner product of an object in a high-dimensional feature space is calculated easily by using kernel function K, and is known as a kernel trick [3, 17, 18].

The traces of the kernelized fuzzy covariance matrix (KW_{tr}) and kernelized Xie-Beni's index (KXB) are introduced herein. In addition, KW_{tr} is constructed by introducing a kernel function into Eq. (5) as follows:

$$KW_{\text{tr}} = \sum_{i=1}^{c} \text{tr}KF_i, \quad \ldots \quad (8)$$

$$\text{tr}KF_i = \frac{1}{U_i} \sum_{k=1}^{n} (u_{ki})^m \|\phi(x_k) - W_i\|^2,$$

$$KF_i = \frac{\sum_{k=1}^{n} (u_{ki})^m (\phi(x_k) - v_i)(\phi(x_k) - v_i)^T}{U_i},$$

$$U_i = \sum_{k=1}^{n} (u_{ki})^m.$$

Moreover, KXB is constructed by introducing a kernel function into Eq. (6) as follows:

$$KXB = \frac{\sum_{k=1}^{n} \sum_{i=1}^{c} (u_{ki})^m \|\phi(x_k) - W_i\|^2}{n \min_{1 \le i,j, \le c,\ i \ne j} \|W_i - W_j\|^2}. \quad \ldots \quad (9)$$

In Eqs. (8) and (9), W_i is the cluster center in $\Re^s$ derived from kernelized fuzzy c-means and is described as follows:

$$W_i = \frac{\sum_{k=1}^{n} (u_{ki})^m \phi(x_k)}{\sum_{k=1}^{n} (u_{ki})^m}. \quad \ldots \quad (10)$$

To calculate the dissimilarity in the high-dimensional feature space, substitute Eq. (10) into $\|\phi(x_k) - W_i\|^2$ and $\|W_i - W_j\|^2$, which are calculated through the following formula:

$$\begin{aligned} d_{ki} &= \|\phi(x_k) - W_i\|^2 \\ &= \langle \phi(x_k), \phi(x_k) \rangle - 2\langle \phi(x_k), W_i \rangle + \langle W_i, W_i \rangle \\ &= K(x_k, x_k) - \frac{2}{U_i} \sum_{s=1}^{n} u_{si} K(x_k, x_s) \\ &\quad + \frac{1}{(U_i)^2} \sum_{s=1}^{n} \sum_{t=1}^{n} u_{si} u_{ti} K(x_s, x_t), \end{aligned}$$

$$\begin{aligned} \|W_i - W_j\|^2 &= \langle W_i, W_i \rangle - 2\langle W_i, W_j \rangle + \langle W_j, W_j \rangle \\ &= \frac{1}{(U_i)^2} \sum_{s=1}^{n} \sum_{t=1}^{n} (u_{si} u_{ti})^m K(x_s, x_t) \\ &\quad - \frac{2}{U_i U_j} \sum_{s=1}^{n} \sum_{t=1}^{n} (u_{si} u_{tj})^m K(x_s, x_t) \\ &\quad + \frac{1}{(U_j)^2} \sum_{s=1}^{n} \sum_{t=1}^{n} (u_{sj} u_{tj})^m K(x_s, x_t). \end{aligned}$$

3. Two-Stage Clustering Based on Cluster Validity Measures

We proposed a two-stage clustering framework based on cluster validity measures [14]. The method is also constructed using a generating stage and merging stage as well as conventional two-stage clustering [7–9]. In the first stage, HCM generates a large number of small clusters, which are called sub-clusters. In the second stage, clusters obtained in the first stage are merged sequentially through the AHC procedure. The difference between the proposed and conventional methods is that the cluster validity measures are used as the dissimilarity in the AHC procedure. The algorithm merges the best cluster pair showing the maximum or minimum value of the cluster validity measures by merging two clusters. The cluster validity measure is used as the criterion to merge or not merge two clusters. The advantage of the proposed method is that the optimal cluster partition is obtained using suitable cluster validity measures as the merging criterion in the second stage. Assume that $CV(G, G')$ is the value of the cluster validity measure when merging two clusters G and G'. Several cluster validity measures [10–13] were applied to the merging criterion in the previous experiments [14]. The proposed method considers the minimum value of the cluster validity measures and finds pairs as follows:

$$(G_p, G_q) = \arg \min_{G, G' \in \mathscr{G}} CV(G, G'). \quad \ldots \quad (11)$$

If we use PC and MPC, the pair that maximizes $CV(G, G')$ is considered.

The algorithm of the two-stage clustering based on the cluster validity measures, abbreviated as HCM–AHC, is summarized as **Algorithm 3**.

In the previous study, we applied cluster validity measures with fuzzy partition as the merging criterion in the second stage. The fuzzy partition is calculated through Eq. (3) based on the centers of the clusters in $\mathscr{G}$ before calculating the cluster validity measures in **STEP 2-1**. After merging the two clusters in **STEP 2-2**, the cluster center of G_r is updated using Eq. (2). In our algorithm, the membership degree for calculating the cluster validity measures is calculated by Eq. (3), and the cluster centers are calculated through Eq. (2). In our previous study, we compared the crisp and fuzzy partitions using artificial and benchmark datasets [14]. The experimental results showed that cluster validity measures based on a fuzzy partition are a suitable merging criterion in the second stage. In this sense, we consider a fuzzy partition to

Algorithm 3 HCM–AHC

STEP 1 Generate clusters using HCM.

STEP 1-1 Set $c := c_0$ and initial cluster centers from X.

STEP 1-2 Calculate $u_{ki} \in U$ using Eq. (1).

STEP 1-3 Calculate $v_i \in V$ using Eq. (2).

STEP 1-4 If the convergence criterion is satisfied go to **STEP 2**. Otherwise, go back to **STEP 1-2**.

STEP 2 Merge clusters through AHC procedure based on cluster validity measures.

STEP 2-1 Assume that the initial clusters are given by **STEP 1**

$$\mathscr{G} = \{G_1, G_2, \dots, G_{c_0}\}$$

Set $C := c_0$ (C is the number of clusters in the AHC procedure, and c_0 is the initial number of clusters).

Calculate $CV(G, G')$ for all pairs $G, G' \in \mathscr{G}$.

STEP 2-2 Search the pair of optimal cluster validity measures using Eq. (11):

Merge: $G_r := G_p \cup G_q$.

Add: G_r to $\mathscr{G}$ and delete G_p and G_q from $\mathscr{G}$.

$C := C - 1$. If $C = 1$, then stop and output the dendrogram.

STEP 2-3 Update $CV(G, G')$ for all pairs $G, G' \in \mathscr{G}$. Go to **STEP 2-2**.

calculate the cluster validity measures with our two-stage clustering algorithm.

4. Experiments

We conducted numerical experiments with eight artificial datasets to verify the effectiveness of the proposed method. Next, we first describe the calculation conditions of the numerical experiments. Second, we describe the clustering results of the two-stage clustering based on the cluster validity measures. Third, we summarize the experimental results and features of the proposed method.

4.1. Experimental Setup

We used eight artificial datasets for the experiments. A description of these datasets is provided in **Table 2**. **Figs. 1–8**, show illustrative examples of artificial datasets 1–8 classified into adequate clusters. To show the effectiveness of the proposed method, we evaluate the value of the adjusted rand index (ARI) [19]. The ARI is a measure of the similarity between two cluster partitions and takes a value between -1 and 1. When the value of ARI is 1, the two cluster partitions are exactly the same.

We execute 100 HCM trials with different initial values

Table 2. Description of eight data sets.

Data	n	p	c
Artificial data 1	300	2	6
Artificial data 2	500	2	4
Artificial data 3	1000	2	3
Artificial data 4	300	2	3
Artificial data 5	500	2	2
Artificial data 6	312	2	2
Artificial data 7	150	2	2
Artificial data 8	150	2	2

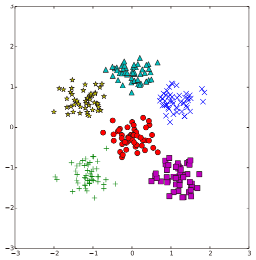


Fig. 1. Artificial dataset 1 ($n = 300$, $p = 2$, $c = 6$).

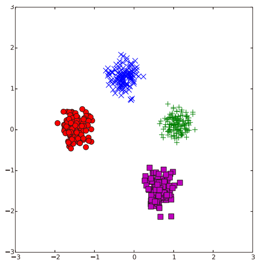


Fig. 2. Artificial dataset 2 ($n = 500$, $p = 2$, $c = 4$).

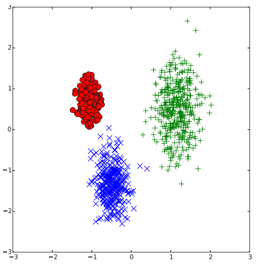


Fig. 3. Artificial dataset 3 ($n = 1000$, $p = 2$, $c = 3$).

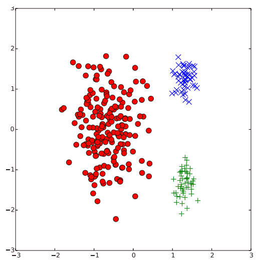


Fig. 4. Artificial dataset 4 ($n = 300$, $p = 2$, $c = 3$).

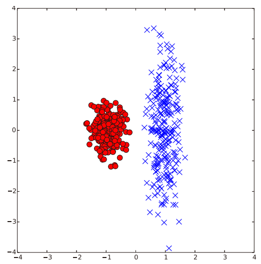


Fig. 5. Artificial dataset 5 ($n = 500$, $p = 2$, $c = 2$).

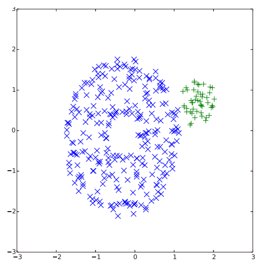


Fig. 6. Artificial dataset 6 ($n = 312$, $p = 2$, $c = 2$).

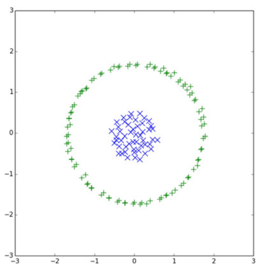


Fig. 7. Artificial dataset 7 ($n = 150$, $p = 2$, $c = 2$).

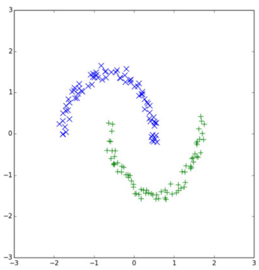


Fig. 8. Artificial dataset 8 ($n = 150$, $p = 2$, $c = 2$).

Table 3. ARI values by the proposed and conventional methods.

	Data 1	Data 2	Data 3	Data 4	Data 5	Data 6	Data 7	Data 8
HCM-AHC(W_{tr})	0.960	1.000	0.994	0.978	1.000	0.112	−0.024	1.000
HCM-AHC(XB)	0.789	0.880	0.622	1.000	1.000	0.029	−0.037	1.000
HCM-AHC(FS)	0.737	0.259	0.136	1.000	0.326	0.846	−0.054	0.099
HCM-AHC(PC)	0.984	1.000	0.989	1.000	1.000	0.846	−0.024	1.000
HCM-AHC(PE)	0.960	1.000	0.994	0.872	1.000	0.846	−0.024	1.000
HCM-AHC(MPC)	0.984	1.000	0.989	1.000	1.000	0.846	−0.024	1.000

during the first stage to minimize the objective function. First, we set the number of clusters generated in the first stage to 10 ($c_0 = 10$) for all datasets. The clusters with the minimum value of the objective function obtained in the first stage are merged in the second stage. We evaluate the value of ARI when the number of clusters determined through the proposed method is equal to the optimal number of clusters. In these experiments, we set $m = 2.0$. To calculate KW_{tr} and KXB, we use the following Gaussian kernel as the kernel function:

$$K(x,y) = \exp\left(-\beta\|x-y\|^2\right). \quad \text{(12)}$$

Here, β is a kernel parameter for the Gaussian kernel. In these experiments, we set β =0.1, 1.0, and 10.0, respectively.

4.2. Experimental Results

We show the experimental results to demonstrate the effectiveness of two-stage clustering based on the cluster validity measures. For each dataset, we evaluate the ARI for clustering accuracy. **Table 3** shows a summary of the results. In **Table 3**, except for HCM-AHC(FS), artificial datasets 7, and 8, the results are from our previous study [14].

The second through sixth rows show the results by the proposed method with each index. **Table 3** shows that several of the indices obtain better results except for artificial dataset 7, which consists of a non-linear boundary. HCM-AHC(W_{tr}) and HCM-AHC(XB) do not obtain better results for artificial dataset 6. None of the two-stage clustering methods based on each measure obtain better results for artificial dataset 7, whereas, with the exception of HCM-AHC(FS), the proposed method can obtain better results for artificial dataset 8, which consists of a non-linear cluster boundary.

To compare the results by HCM-AHC(W_{tr}) and HCM-AHC(XB), we then conducted experiments using HCM-AHC(KW_{tr}) and HCM-AHC(KXB) for artificial data 6–8. **Tables 4–6** show the results by HCM-AHC(KW_{tr}) and HCM-AHC(KXB) for artificial datasets 6, 7, and 8. The bold face in **Tables 4–6** shows the best results of each comparison.

A summary of the comparison is as follows:

- The kernelized indices with $\beta = 0.1$ obtain similar results for artificial dataset 6.

Table 4. ARI value by kernel indices for artificial dataset 6.

	$\beta = 0.1$	$\beta = 1.0$	$\beta = 10.0$
HCM-AHC(KW_{tr})	**0.846**	0.187	0.806
HCM-AHC(KXB)	**0.846**	0.289	−0.103

Table 5. ARI value by kernel indices for artificial dataset 7.

	$\beta = 0.1$	$\beta = 1.0$	$\beta = 10.0$
HCM-AHC(KW_{tr})	−0.046	−0.037	**1.000**
HCM-AHC(KXB)	−0.048	−0.048	−0.023

Table 6. ARI value by kernel indices for artificial dataset 8.

	$\beta = 0.1$	$\beta = 1.0$	$\beta = 10.0$
HCM-AHC(KW_{tr})	0.016	0.105	0.008
HCM-AHC(KXB)	0.251	0.406	**0.749**

- HCM-AHC(KW_{tr}) with $\beta = 10.0$ shows the best results for artificial dataset 7.
- The kernelized indices do not obtain better results for artificial dataset 8 because they have a possibility to merge clusters that are distant from each other. Conventional indices such as W_{tr}, and XB, and partition coefficients are considered suitable merging criteria for non-linearity, such as with artificial dataset 8.

4.3. Discussions

First, we show the overview of indices. **Tables 3–6** show that several of the cluster validity measures are suitable for a merging criterion in the second stage.

- HCM-AHC(W_{tr}) obtains better results except for artificial datasets 6 and 7. If we set the number of clusters generated in the first stage to 20 ($c_0 = 20$), we can obtain the best result when ARI = 1.0 for artificial dataset 6. For artificial dataset 7, **Table 5** shows that KW_{tr} obtains better results than the other indices.
- HCM-AHC(XB) obtains relatively poor results compared with the other indices for artificial datasets 1, 2, 3, and 6. The XB index (shown in Eq. (6)) consists of the average of the sum of squares within the clusters and the minimum dissimilarity of two cluster centers. Therefore, there is a possibility for clusters

that are distant from each other to be merged when the denominator of Eq. (6) significantly affects *XB*. The examples and a detailed discussion are described in our previous study [14].

- HCM-AHC(*FS*) obtains more negative results than the other indices. The value of *FS* is strongly affected by the second term in Eq. (7). This means that the *FS* is not suitable for the criterion in the second stage, which merges adjacent clusters. HCM-AHC(*FS*) with $c_0 = 20$ also obtains negative results for artificial datasets 2, 3, and 5. These results indicate that *FS* is not suitable for the merging criterion in the second stage.
- It can be seen that the indices *PC*, *PE*, and *MPC*, which are based on the membership degree, obtain better results even for artificial datasets 6 and 8, which consist of different cluster shapes.
- **Tables 4–6** show that KW_{tr} is more suitable for the merging criterion than *KXB*. It is considered that KW_{tr} is as suitable for merging adjacent clusters as W_{tr}.

These experiments show that W_{tr}, *PC*, *PE*, and *MPC* are suitable for the merging criterion for two-stage clustering based on the cluster validity measures. In particular, it is considered that W_{tr} and its kernelization KW_{tr} are the best indices for the proposed method. The strong point of the proposed method is the handling of complex structures by considering the cluster validity measures as the merging criterion. To improve the effectiveness of the proposed method, the introduction of fast algorithms for speeding-up techniques in each stage are important for handling massive complex datasets.

5. Conclusions

In this paper, we proposed a two-stage clustering method based on cluster validity measures. The proposed method is different from previous studies [7–9] in that it handles cluster validity measures as the merging criterion in the second stage. Moreover, we demonstrated the effectiveness and characteristics of each cluster validity measure through numerical experiments on eight datasets. These experiments show that W_{tr} and its kernelization KW_{tr} are quite effective for two-stage clustering based on cluster validity measures.

In future works, to accelerate each stage, we will apply other clustering methods in the first stage and novel cluster validity measures in the second stage. We will also propose and apply variants and other cluster validity measures to improve the proposed method and handle massive and complex datasets.

Acknowledgements

This work was partly supported by JSPS KAKENHI Grant Numbers JP16K16128 and JP17K00332. This work was also supported in part by Kindai University Research Enhancement Grant, Faculty Assistance and Development Research Grant Number SR01.

References:

[1] A. K. Jain, "Data clustering: 50 years beyond *K*-means," Pattern Recognition Letters, Vol.31, No.8, pp. 651-666, 2010.
[2] J. C. Bezdek, "Pattern Recognition with Fuzzy Objective Function Algorithms," Plenum Press, New York, 1981.
[3] S. Miyamoto, H. Ichihashi, and K. Honda, "Algorithms for Fuzzy Clustering," Springer, Heidelberg, 2008.
[4] S. Miyamoto, "Introduction to Cluster Analysis," Morikita-shuppan, 1999 (in Japanese).
[5] G. Hamerly, "Making *k*-means even faster," Proc. of the 2010 SIAM Int. Conf. on Data Mining (SDM), pp. 130-140, 2010.
[6] T. C. Havens, J. C. Bezdek, C. Leckie, L. O. Hall, and M. Palaniswami, "Fuzzy *c*-means algorithms for very large data," IEEE Trans. on Fuzzy Systems, Vol.20, No.6, pp. 1130-1146, 2012.
[7] N. Obara and S. Miyamoto, "A method of two-stage clustering with constraints using agglomerative hierarchical algorithm and one-pass *k*-means," Proc. of The 6th Int. Conf. on Soft Computing and Intelligent Systems The 13th Int. Symposium on Advanced Intelligent Systems (SCIS&ISIS2012), pp. 1540-1544, 2012.
[8] Y. Tamura and S. Miyamoto, "Two-stage clustering using one-pass *k*-medoids and medoid-based agglomerative hierarchical algorithms," Proc. of The 7th Int. Conf. on Soft Computing and Intelligent Systems The 15th Int. Symposium on Advanced Intelligent Systems (SCIS&ISIS2014), pp. 484-488, 2014.
[9] Y. Tamura and S. Miyamoto, "A method of two stage clustering using agglomerative hierarchical algorithms with one-pass *k*-means++ or *k*-median++," Proc. of The 2014 IEEE Int. Conf. on Granular Computing (GrC2014), pp. 281-285, 2014.
[10] D. L. Davies and D. W. Bouldin, "A cluster separation measure," IEEE Trans. on Pattern Analysis and Machine Intelligence, Vol.1, No.2, pp 224-227, 1979.
[11] I. Gath and A. B. Geva, "Unsupervised optimal fuzzy clustering," IEEE Trans. on Pattern Analysis and Machine Intelligence, Vol.11, No.7, pp. 773-780, 1989.
[12] X. L. Xie and G. Beni, "A validity measure for fuzzy clustering," IEEE Trans. on Pattern Analysis and Machine Intelligence, Vol.13, No.8, pp. 841-847, 1991.
[13] W. Wang and Y. Zhang, "On fuzzy cluster validity indices," Fuzzy Sets and Systems, Vol.158, No.19, pp. 2095-2117, 2007.
[14] R. Ozaki, Y. Hamasuna, and Y. Endo, "A method of two-stage clustering based on cluster validity measures," Proc. of The 8th Int. Conf. on Soft Computing and Intelligent Systems The 17th Int. Symposium on Advanced Intelligent Systems (SCIS&ISIS2016), pp. 410-415, 2016.
[15] W. Hashimoto, T. Nakamura, and S. Miyamoto, "Comparison and evaluation of different cluster validity measures including their kernelization," J. Adv. Comput. Intell. Intell. Inform., Vol.13, No.3, pp. 204-209, 2009.
[16] Y. Fukuyama and M. Sugeno, "A new method of choosing the number of clusters for fuzzy *c*-means method," Proc. of 5th Fuzzy system Symposium, , pp. 247-250, 1989.
[17] M. Girolami, "Mercer kernel-based clustering in feature space," IEEE Trans. Neural networks, Vol.13, No.3, pp. 780-784, 2002.
[18] Y. Endo, H. Haruyama, and T. Okubo, "On some hierarchical clustering algorithms using kernel functions, " Proc. of IEEE Int. Conf. on Fuzzy Systems(FUZZ-IEEE2004), pp. 1513-1518, 2004.
[19] L. Hubert and P. Arabie, "Comparing Partitions," J. of Classification, Vol.2, No.1, pp. 193-218, 1985.

Name:
Yukihiro Hamasuna

Affiliation:
Lecturer, Department of Informatics, School of Science and Engineering, Kindai University

Address:
3-4-1 Kowakae, Higashiosaka, Osaka 577-8502, Japan

Brief Biographical History:
2009-2010 Research Fellow of the Japan Society for the Promotion of Science
2011-2014 Assistant Professor, Kindai University
2014- Lecturer, Kindai University

Main Works:
• Y. Hamasuna, N. Kinoshita, and Y. Endo, "Comparison of Cluster Validity Measures Based *x*-means," J. Adv. Comput. Intell. Intell. Inform., Vol.20, No.5, pp. 845-853, 2016.
• Y. Hamasuna and Y. Endo, "On Sequential Cluster Extraction Based on L_1-Regularized Possibilistic *c*-Means," J. Adv. Comput. Intell. Intell. Inform., Vol.19, No.5, pp. 655-661, 2015.
• Y. Hamasuna and Y. Endo, "On Semi-supervised Fuzzy *c*-Means Clustering for Data with Clusterwise Tolerance by Opposite Criteria," Soft Computing, Vol.17, No.1, pp. 71-81, 2013.
• Y. Hamasuna, Y. Endo, and S. Miyamoto, "On Tolerant Fuzzy *c*-Means Clustering and Tolerant Possibilistic Clustering," Soft Computing, Vol.14, No.5, pp. 487-494, 2010.

Membership in Academic Societies:
• Japan Society for Fuzzy Theory and Systems (SOFT)
• The Japanese Society for Artificial Intelligence (JSAI)
• The Institute of Electrical and Electronics Engineering (IEEE)

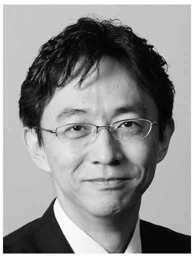
Name:
Ryo Ozaki

Affiliation:
Graduate School of Science and Engineering, Kindai University

Address:
3-4-1 Kowakae, Higashiosaka, Osaka 577-8502, Japan

Brief Biographical History:
2016- Graduate School of Science and Engineering, Kindai University

Main Works:
• R. Ozaki, Y. Hamasuna, and Y. Endo, "Agglomerative Hierarchical Clustering Based on Local Optimization for Cluster Validity Measures," Proc. of 2017 IEEE Int. Conf. on Systems, Man, and Cybernetics (IEEE SMC 2017), pp. 1822-1827, 2017.
• R. Ozaki, Y. Hamasuna, and Y. Endo, "A Method of Two-Stage Clustering Based on Cluster Validity Measures," Joint 8th Int. Conf. on Soft Computing and Intelligent Systems and 17th Int. Symp. on Advanced Intelligent Systems (SCIS & ISIS 2016), pp. 410-415, 2016.



Name:
Yasunori Endo

Affiliation:
Professor, Faculty of Engineering, Information and Systems, University of Tsukuba

Address:
1-1-1 Tennodai, Tsukuba, Ibaraki 305-8573, Japan

Brief Biographical History:
1994-1997 Research Assistant, Waseda University
1997-2001 Assistant Professor, Tokai University
2001-2004 Assistant Professor, University of Tsukuba
2004-2013 Associate Professor, University of Tsukuba
2012 Guest Research Scholar, International Institute for Applied Systems Analysis (IIASA)
2013- Professor, University of Tsukuba

Main Works:
• Y. Endo and S. Miyamoto, "Spherical k-Means++ Clustering," The 12th Int. Conf. on Modeling Decisions for Artificial Intelligence (MDAI 2015), Springer, LNAI 9321, pp. 103-114, 2015.
• Y. Hamasuna and Y. Endo, "On a Family of New Sequential Hard Clustering," J. Adv. Comput. Intell. Intell. Inform., Vol.19, No.6, pp. 759-765, 2015.
• Y. Endo and N. Kinoshita, "Objective-Based Rough *c*-Means Clustering," Int. J. of Intelligent Systems, Vol.28, Issue 9, pp. 907-925, 2013.
• Y. Endo, Y. Hasegawa, Y. Hamasuna, and S. Miyamoto, "Fuzzy *c*-Means Clustering for Uncertain Data using Quadratic Penalty-Vector Regularization," J. Adv. Comput. Intell. Intell. Inform., Vol.15, No.1, pp. 76-82, 2011.
• Y. Endo, R. Murata, H. Haruyama, and S. Miyamoto, "Fuzzy *c*-Means for Data with Tolerance," Proc. 2005 Int. Symp. on Nonlinear Theory and Its Applications, pp. 345-348, 2005.
• Y. Endo, H. Haruyama, and T. Okubo, "On Some Hierarchical Clustering Algorithms Using Kernel Functions," Proc. IEEE Int. Conf. on Fuzzy Systems, 2004.

Membership in Academic Societies:
• Japan Society for Fuzzy Theory and Intelligent Informatics (SOFT)
• The Institute of Electronics, Information and Communication Engineers (IEICE)
• The Institute of Electrical and Electronics Engineers (IEEE)