A Clustering Criterion Based on Distortion Ratios and Its Algorithms

Fujiki Morii

***Abstract*—A clustering criterion based on distortion ratios and its algorithms are proposed without offering the knowledge of the number of clusters. Computing distortion ratios on splitting and distortion ratios on merging for clusters of a data set, the criterion function is defined as the mean of the Euclidian distances between points of those distortion ratios and a reference point. Three algorithms are proposed, whose algorithms are designed to optimize the criterion function over the number of clusters and partitions of the data set. Through several classification experiments, the effectiveness of the criterion and those algorithms is demonstrated.**

***Index Terms*—clustering, criterion, distortion ratio, algorithm, number of clusters.**

I. INTRODUCTION

CLUSTERING [1-7], whose aim is to classify an unlabeled data set to appropriate clusters, is an important and fundamental research issue for pattern recognition, image processing and data mining. The ideal of clustering is to find the proper number of clusters and obtain a good partition of the data set without external information [8-16].

The main purpose of this paper is to propose a new clustering criterion based on distortion ratios and three algorithms to realize the criterion under no external knowledge of the proper number of clusters [17]. Executing split operations and merge operations tentatively for clusters obtained by partitioning the data set, a distortion ratio on splitting and a distortion ratio on merging for each cluster are computed, whose ratios provide measures of unimodality or nonunimodality of the cluster [1][16].

The criterion function is defined as the mean of the Euclidian distances between points of those distortion ratios and a reference point, whose reference point expresses a standard point of the distortion ratios which a cluster should neither be split nor be merged. The function is optimized by minimizing over the number of clusters and clusters, i.e., partitions of the data set. By executing split and merge operations for clusters, this criterion is designed for each cluster to maintain unimodality.

To realize the criterion, three algorithms are proposed. Concerning the first algorithm, starting with a set of initial clusters, a splitting operation or a merging operation is iterated to minimize the function over the clusters until finding a local minimum. Selecting several random sets of initial clusters, this procedure is iterated, and we obtain a set of clusters minimizing the function over classification results by those procedures. The second algorithm modifies the first one to decrease the computation time. Minimizing the criterion function by using k-means algorithm (KMA) [7] for the variation of the number of clusters, the third algorithm selects the optimum number of clusters and a classification result, .

Through several classification experiments, the effectiveness of the clustering criterion and those algorithms is demonstrated.

II. CLUSTERING CRITERION

Let us treat a data set X composed of n samples $\vec{x}_i = (x_{i1}, \ldots, x_{iU}), i = 1, \ldots, n$, where samples are U-dimensional real vectors and the proper number of clusters in X is unknown. By classifying X without supervision, X is partitioned into k disjoint subsets $X_l, l = 1, \ldots, k$, whose subsets are called clusters.

Our research purpose is to propose a clustering criterion and its algorithms to obtain the proper number of clusters k and an appropriate set of clusters for the data set.

Let us assume that a set of clusters $P = \{X_l, l = 1, \ldots, k\}$ is provided. By executing a split operation and merge operations tentatively for each cluster X_l, a criterion function based on distortion ratios is derived in the following.

The minimum distortion for X_l is defined by

$$D_l(1) = \min_{\vec{c}_l} \sum_{\vec{x}_i \in X_l} d(\vec{x}_i, \vec{c}_l), \tag{1}$$

where $d(\vec{x}_i, \vec{c}_l)$ is a nonnegative distortion measure between $\vec{x}_i$ and $\vec{c}_l = (c_{l1}, \ldots, c_{lU})$. $\vec{c}_l$ satisfying (1) is called the cluster center of X_l and expresses a representative of X_l. As $d(\cdot, \cdot)$, we can use the squared Euclidean distance by

$$\begin{aligned} d(\vec{x}_i, \vec{c}_l) &= \|\vec{x}_i - \vec{c}_l\|^2, \\ &= \sum_{u=1}^{U} (x_{iu} - c_{lu})^2, \end{aligned} \tag{2}$$

or the more general quadratic distortion by

$$d(\vec{x}_i, \vec{c}_l) = (\vec{x}_i - \vec{c}_l) B (\vec{x}_i - \vec{c}_l)^t, \tag{3}$$

where B is a $U \times U$ positive definite symmetric matrix. Other distortion measures can be also exploited in this criterion.

Each cluster X_l is tentatively partitioned into 2 subclusters X_{l1} and X_{l2} to realize the minimum distortion given by

$$D_l(2) = \min_{\{\vec{c}_{lq}\}} \min_{\{X_{lq}\}} \sum_{q=1}^{2} \sum_{\vec{x}_i \in X_{lq}} d(\vec{x}_i, \vec{c}_{lq}), \tag{4}$$

F. Morii is with the Department of Information and Computer Sciences, Nara Women's University, Nara, 630-8506 Japan (e-mail: fmorii@ics.nara-wu.ac.jp).

Manuscript received February 8, 2012; revised xx, 2012.

DOI: 10.5176_2010-2283_2.1.127

where $\{\vec{c}_{lq}, q = 1, 2\}$ are called the cluster centers for X_{l1} and X_{l2}.

Let us introduce a splitting measure for X_l given by

$$S_l = D_l(2)/D_l(1), \tag{5}$$

which is called the distortion ratio on splitting. The small value of S_l states that X_l should be split into 2 subclusters because of low unimodality of X_l. The large value of S_l states that X_l should not be split into 2 subclusters because of high unimodality of X_l.

Next, merge operations for X_l are executed tentatively. For neighboring X_v and X_w, the distortion ratio on merging

$$M_{vw} = (D_v(1) + D_w(1))/D_{vw}(1) \tag{6}$$

is introduced, where $D_v(1)$ and $D_w(1)$ are used according to (1). The distortion of the merged cluster $X_v \cup X_w$ is given by

$$D_{vw}(1) = \min_{\vec{c}_{vw}} \sum_{\vec{x}_i \in X_v \cup X_w} d(\vec{x}_i, \vec{c}_{vw}), \tag{7}$$

where the optimized $\vec{c}_{vw}$ is called the cluster center of $X_v \cup X_w$.

As the distortion ratio on merging for X_l, we define

$$\begin{aligned} M_l &= \max_{w,(w \neq l)} M_{lw} \text{ for all neighboring } X_l, X_w \quad (8)\\ &= M_{l\hat{w}}. \quad (9) \end{aligned}$$

The small value of M_l states that X_l should not be merged because of low unimodality of $X_l \cup X_{\hat{w}}$. The large value of M_l states that X_l and $X_{\hat{w}}$ should be merged because of high unimodality of $X_l \cup X_{\hat{w}}$.

Introducing a two dimensional orthogonal coordinate system where the x-axis expresses the distortion ratio on splitting and the y-axis expresses the distortion ratio on merging, k points $\{(S_l, M_l), l = 1, \cdots, k\}$ corresponding to k clusters are plotted on the plane. We can also show $0 \leq S_l \leq 1$ and $0 \leq M_l \leq 1$. The criterion function based on distortion ratios is defined as the arithmetic mean of the Euclidian distances between those k points and a reference point (x_r, y_r). That is to say, the criterion function is provided by

$$\begin{aligned} R(k, P) &\equiv R(k, \{X_l, l = 1, \cdots, k\}) \\ &= \frac{1}{k} \sum_{l=1}^{k} \sqrt{(S_l - x_r)^2 + (M_l - y_r)^2}, \quad (10) \end{aligned}$$

and the optimized k_{opt} and $P_{opt} \equiv \{X_{l,opt}, l = 1, \cdots, k_{opt}\}$ are determined as k and $\{X_l\}$ minimizing $R(k, \{X_l\})$.

In (10), the reference point (x_r, y_r) expresses a standard point of the distortion ratios whose cluster should neither be split nor be merged, where we usually use $(x_r, y_r) = (1, 0)$. When a cluster $X_\beta = \{\vec{x}_1, \vec{x}_1, \vec{x}_2, \vec{x}_2, \cdots, \vec{x}_\gamma, \vec{x}_\gamma\}$ is split into two subclusters $X_{\beta 1} = X_{\beta 2} = \{\vec{x}_1, \vec{x}_2, \cdots, \vec{x}_\gamma\}$, we have $S_\beta = 1$. In $M_{\beta\hat{w}}$ of (9), when $X_{\hat{w}}$ is infinitely far away from X_β, we obtain $M_\beta \to 0$, which defines $M_\beta = 0$. Hence, when a cluster has $(S_l, M_l) = (1, 0)$, the cluster should neither be split nor be merged definitely. Since the above $S_\beta = 1$ is in a singular situation, we can use a value in $0.6 \leq x_r \leq 1$ practically.

Another version different from the function R is provided by

$$J(k, P) = \frac{1}{k} \sum_{l=1}^{k} \sqrt{(S_l - \tilde{x}_r)^2 + (M_l - \tilde{y}_r)^2}, \tag{11}$$

where $(\tilde{x}_r, \tilde{y}_r) = (0, 1)$. The optimized k_{opt} and P_{opt} are determined as k and P maximizing $J(k, P)$. When a cluster has $(S_l, M_l) = (0, 1)$, the cluster should both be split and be merged definitely.

III. Clustering algorithms

Let us propose three clustering algorithms to minimize the criterion function $R(k, \{X_l\})$ of (10) over k and $\{X_l\}$.

The first algorithm, which we call Algorithm 1, is considered. When time is t, assume that we have a number of clusters $k^{(t)}$ and a set of clusters $P^{(t)} = \{X_l^{(t)}, l = 1, \ldots, k^{(t)}\}$. Then $R(k^{(t)}, P^{(t)})$ is computed by (10).

The next step is to execute a split operation or a merge operation of a cluster to make the function as small as possible.

When splitting $X_l^{(t)}$ into two subclusters by (4), we have $k^{(t+1)} \leftarrow k^{(t)} + 1$ and $P^{(t+1)}$, and compute

$$R_{split}^{(t+1)} \equiv \min_{l=1,\cdots,k^{(t)}} R(k^{(t+1)}, P^{(t+1)} | X_l^{(t)} \text{is split}). \tag{12}$$

When merging $X_v^{(t)}$ and $X_w^{(t)}$ by (6), we have $k^{(t+1)} \leftarrow k^{(t)} - 1$ and $P^{(t+1)}$, and compute

$$R_{merge}^{(t+1)} \equiv \min_{v,w} R(k^{(t+1)}, P^{(t+1)} | X_v^{(t)} \text{and} X_w^{(t)} \text{are merged}). \tag{13}$$

If

$$R_{split}^{(t+1)} = \min\{R(k^{(t)}, P^{(t)}), R_{split}^{(t+1)}, R_{merge}^{(t+1)}\}, \tag{14}$$

the same procedure is iterated after splitting. If

$$R_{merge}^{(t+1)} = \min\{R(k^{(t)}, P^{(t)}), R_{split}^{(t+1)}, R_{merge}^{(t+1)}\}, \tag{15}$$

the same procedure is iterated after merging. If

$$R(k^{(t)}, P^{(t)}) = \min\{R(k^{(t)}, P^{(t)}), R_{split}^{(t+1)}, R_{merge}^{(t+1)}\}, \tag{16}$$

the computation is terminated. Since we have the possibility of obtaining a local minimum when the termination of computation, multiple selections of initial parameters $k^{(0)}$ and $P^{(0)}$ are needed, where $P^{(0)}$ is usually provided by a certain effective clustering algorithm such as KMA or vector quantization (VQ) [6].

The algorithm to optimize $R(k, P)$ is summarized as follows.

(Clustering Algorithm 1 Based on Distortion Ratios)
(CA1) For a given M, repeat (CA2)-(CA6) for $m = 1, \cdots, M$.
(CA2) Set the number of clusters $k^{(0)}$ at random.
(CA3) Obtain a set of initial clusters $P^{(0)}$ by a clustering method such as KMA or VQ with the number of clusters $k^{(0)}$.
(CA4) Repeat (CA5) for $t = 0, 1, \cdots$.
(CA5) If $R_{split}^{(t+1)} = \min\{R(k^{(t)}, P^{(t)}), R_{split}^{(t+1)}, R_{merge}^{(t+1)}\}$, we have $k^{(t+1)} \leftarrow k^{(t)} + 1$ and $P^{(t+1)}$.
If $R_{merge}^{(t+1)} = \min\{R(k^{(t)}, P^{(t)}), R_{split}^{(t+1)}, R_{merge}^{(t+1)}\}$, we have $k^{(t+1)} \leftarrow k^{(t)} - 1$ and $P^{(t+1)}$.

If $R(k^{(t)}, P^{(t)}) = \min\{R(k^{(t)}, P^{(t)}), R^{(t+1)}_{split}, R^{(t+1)}_{merge}\}$, go to (CA6).
(CA6) Store $\{m, k^{(t)}, P^{(t)}, R(k^{(t)}, P^{(t)})\}$.
(CA7) Determine an optimum solution by selecting the minimum of $\{R(k^{(t)}, P^{(t)})\}$.
(End of CA)

Let us consider the second algorithm, which is called Algorithm 2. In Algorithm 1, the computation of minimization in (12) and (13) needs a lot of time. To decrease the time, we select $X^{(t)}_\gamma$ satisfying

$$\gamma = \arg \max_{l=1,\cdots,k^{(t)}} \sqrt{(S_l - x_r)^2 + (M_l - y_r)^2} \tag{17}$$

for $P^{(t)} = \{X^{(t)}_l, l = 1, \ldots, k^{(t)}\}$, where $X^{(t)}_\gamma$ is the most possible cluster of splitting or merging. When splitting $X^{(t)}_\gamma$, we use

$$R^{(t+1)}_{split} \equiv R(k^{(t+1)}, P^{(t+1)} | X^{(t)}_\gamma \text{is split}) \tag{18}$$

instead of (12). When merging $X^{(t)}_\gamma$, we use

$$R^{(t+1)}_{merge} \equiv \min_w R(k^{(t+1)}, P^{(t+1)} | X^{(t)}_\gamma \text{and} X^{(t)}_w \text{are merged}) \tag{19}$$

instead of (13). Algorithm 2 is one which (18) and (19) are used in Algorithm 1.

Next, let us consider Algorithm 3, where the criterion $R(k, P)$ of (10) is executed on the classified data by KMA having each number of clusters k for $k_{\min} \leq k \leq k_{\max}$. Algorithm 3 is provided as follows.

(Clustering Algorithm 3 Based on Distortion Ratios)
(CA3-1) For a region of the number of clusters $k_{\min} \leq k \leq k_{\max}$, repeat (CA3-2)-(CA3-3).
(CA3-2) Classify the data set by KMA having the number of clusters k. We obtain clusters $P(k) = \{X^{(k)}_1, \cdots, X^{(k)}_k\}$.
(CA3-3) Compute $R(k, P(k))$ and store it.
(CA3-4) Determine the optimum k_{opt} and clusters $P_{opt}(k_{opt})$ taking $\min\{R(k, P(k)), k_{\min} \leq k \leq k_{\max}\}$.
(End of CA3)

IV. CLASSIFICATION EXPERIMENTS

Let us consider a data set "$k5$" having 5 classes shown by Fig.1, where correct clusters are called classes. The information of the data set $k5$ is provided by Table 1. For classification experiments, we use the squared Euclidian distance as $d(\cdot, \cdot)$ and use KMA to obtain a set of initial clusters $P^{(0)}$ and $D_l(2)$ of (4). For the computation of R by (10), we also adopt $(x_r, y_r) = (1, 0)$.

TABLE I
DATA INFORMATION ON THE DATA SET $k5$.

Class	Number	Centroid	Variance	Covariance
Class 1	60	(0.117,0.265)	(-0.027,0.346)	-0.028
Class 2	45	(1.49,0.247)	(7.09,0.150)	-0.038
Class 3	40	(4.07,0.166)	(4.0,0.201)	-0.035
Class 4	85	(4.68,1.43)	(-4.04,1.18)	0.342
Class 5	100	(9.98,1.69)	(2.91,1.52)	0.0022

In Algorithm 1, let us begin with an initial number of clusters $k^{(0)} = 7$. We have the classification result shown in

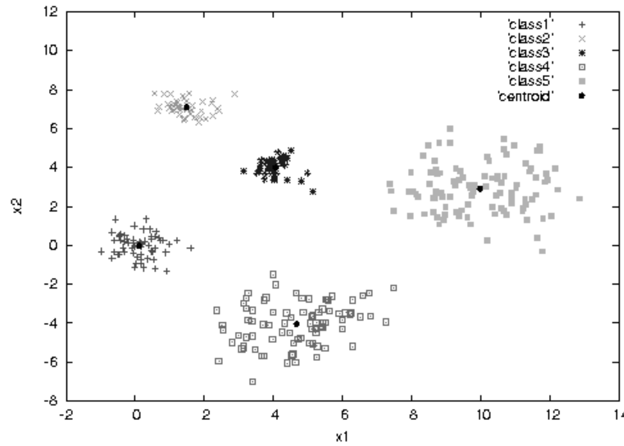

Fig. 1. The data set $k5$ composed of 5 classes.

Fig. 2 by applying KMA with $k^{(0)} = 7$ to the data set $k5$. In the set of initial clusters $P^{(0)} = \{X1, \cdots, X7\}$ of Fig.2, it is recognized that class4 is split into $X4$ and $X5$ and class5 is split into $X6$ and $X7$.

Fig.3 shows $\{(S_l, M_l)\}$ for the clusters in $P^{(0)}$. Concerning to $\{X4, X5, X6, X7\}$, the distortion ratios on splitting have moderate values and the distortion ratios on merging have moderately large values. Hence, $\{X4, X5, X6, X7\}$ should not be split but be merged.

Concerning to $\{X1, X2, X3\}$, since the distortion ratios on splitting have moderate values and the distortion ratios on merging are small, $\{X1, X2, X3\}$ should neither be split nor be merged.

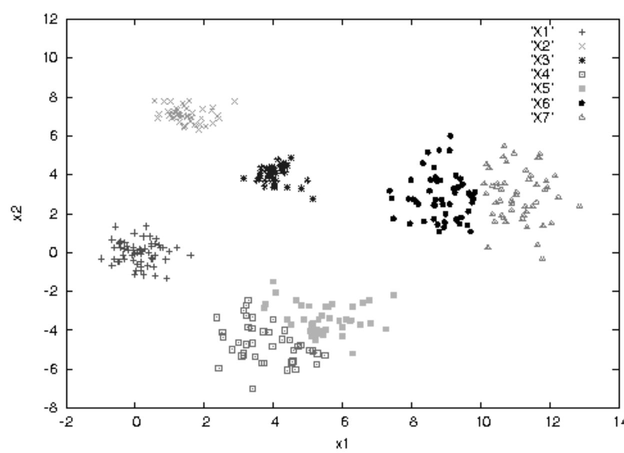

Fig. 2. A set of initial clusters $P^{(0)}$ for the data set $k5$ by using KMA with $k^{(0)} = 7$.

To decrease the criterion function R, $X7$ is merged with $X6$, and we have $P^{(1)} = \{X1, X2, X3, X4, X5, X6 \cup X7\}$. Furthermore, to decrease the criterion function R, $X4$ is merged with $X5$. $P^{(2)} = \{X1, X2, X3, X4 \cup X5, X6 \cup X7\}$ is obtained. The set of clusters $P^{(2)}$ is the optimized one minimizing the function R, where the classification result becomes Fig.4 and we rename $Xl \leftarrow Xl (l = 1, 2, 3)$, $X4 \leftarrow (X4 \cup X5)$ and $X5 \leftarrow (X6 \cup X7)$.

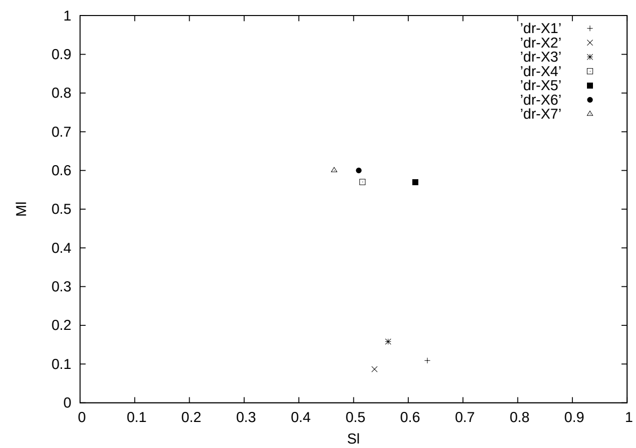

Fig. 3. Distortion ratios on splitting and merging for 7 clusters $X_1, \cdots, X_7$ in Fig.2.

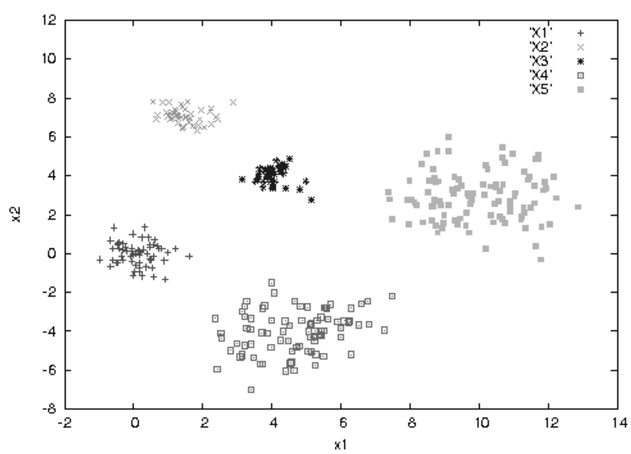

Fig. 4. The data set $k5$ split into 5 clusters $X1, \cdots, X5$.

Fig.5 shows $\{(S_l, M_l)\}$ for clusters in $P^{(2)}$ of Fig.4. Concerning to $\{X1, \cdots, X5\}$, since the distortion ratios on splitting have moderate values and the distortion ratios on merging are small, $\{X1, \cdots, X5\}$ should neither be split nor be merged.

In this case, the optimal classification result can be obtained by the criterion function R with an initial value $k^{(0)} = 7$. We usually use multiple values of $k^{(0)}$ to avoid a local minimum of the function.

Fig.6 provides the performance on the number of clusters k vs. the criterion function $R(k, P)$ for the data set $k5$ and the reference point $(x_r, y_r) = (1, 0)$ when using Algorithm 3. We obtain $k_{opt} = 5$ as the optimum number of clusters. Generally speaking, classification results by Algorithms 1-3 do not necessarily make no difference. In this case we have the same results.

Let us treat another data set "$k7$" shown by Fig.7, whose set is composed of 7 classes and has the data information given by Table 2.

We obtain the performance on the number of clusters k vs. the criterion function $R(k, P)$ for the data set $k7$ and the reference points $(x_r, y_r) = (1, 0), (0.8, 0), (0.6, 0)$ when using

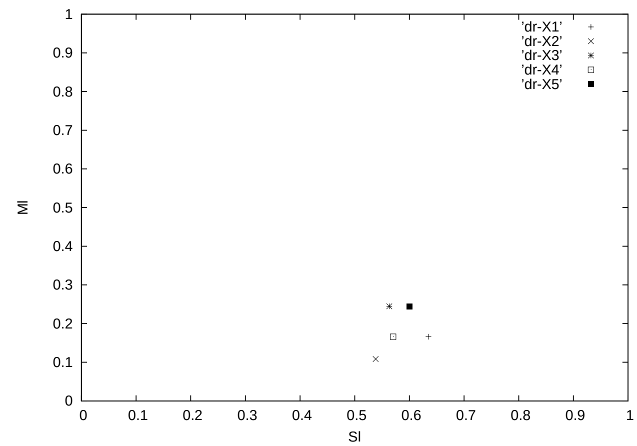

Fig. 5. Distortion ratios on splitting and merging for 5 clusters $X_1, \cdots, X_5$.

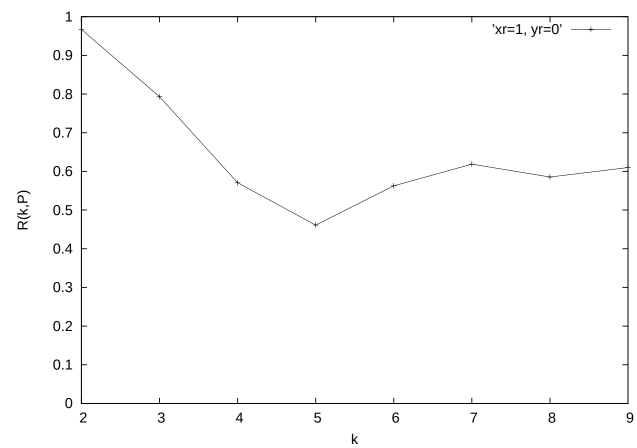

Fig. 6. Number of clusters k vs. criterion function $R(k, P)$ for the data set $k5$ and the reference points $(x_r, y_r) = (1, 0)$ when using Algorithm 3.

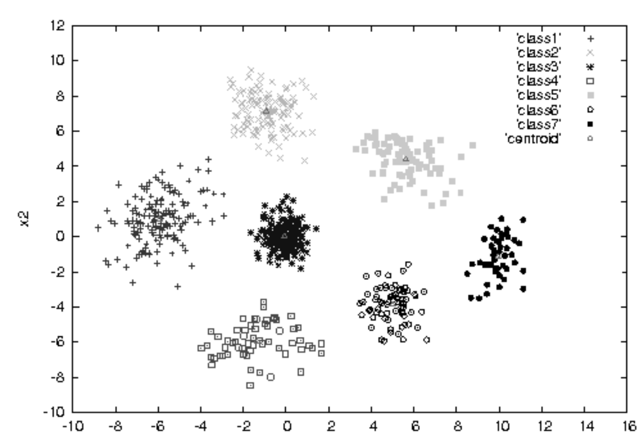

Fig. 7. The data set $k7$ composed of 7 classes.

TABLE II
DATA INFORMATION ON THE DATA SET $k7$.

Class	Number	Centroid	Variance	Covariance
Class 1	150	(-5.94,0.994)	(1.26,1.83)	0.492
Class 2	115	(-0.919,7.14)	(0.828,1.14)	-0.142
Class 3	200	(-0.070,0.025)	(0.289,0.529)	-0.0014
Class 4	60	(-1.35,-5.98)	(1.93,0.949)	0.197
Class 5	79	(5.61,4.39)	(1.28,1.03)	-0.399
Class 6	69	(4.99,-3.87)	(0.520,0.965)	-0.032
Class 7	45	(9.94,-1.15)	(0.372,1.36)	0.239

Algorithm 3, whose performance is provided by Fig.8. From Fig.8, we can acquire the optimum number of clusters $k_{opt} = 7$. We can also obtain the same results by using Algorithm 1 and Algorithm 2.

The clustering result by KMA with $k_{opt} = 7$ is given by Fig.9, where two samples are misclassified.

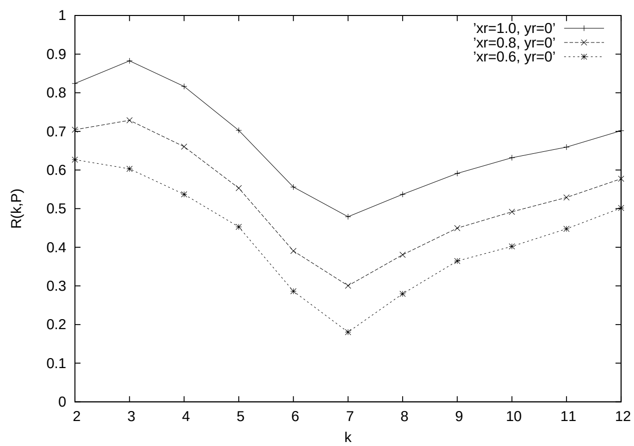

Fig. 8. Number of clusters k vs. criterion function $R(k, P)$ for the data set $k7$ and the reference points $(x_r, y_r) = (1, 0), (0.8, 0), (0.6, 0)$ when using Algorithm 3.

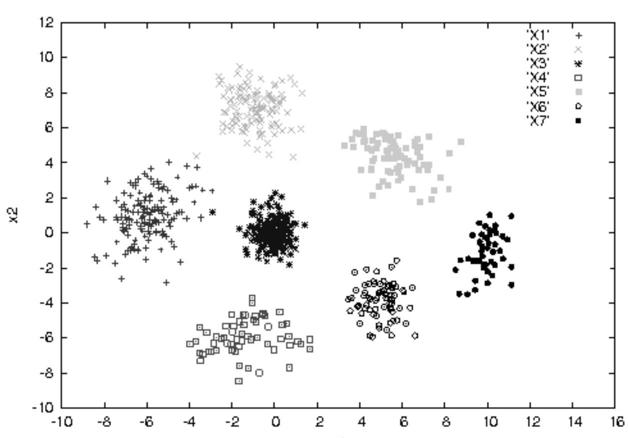

Fig. 9. The data set $k7$ split into 7 clusters $X_1, \cdots, X_7$ by KMA with $k = 7$.

V. CONCLUSION

A clustering criterion and its algorithms for a data set without offering the knowledge of the number of clusters were introduced. Its criterion function is composed of distortion ratios on splitting and distortion ratios on merging for clusters, and to minimize the function means attaining the situation that each cluster should neither be split nor be merged. Those algorithms realizing the clustering criterion are organized to minimize the function by executing split operations and merge operations. Using the criterion and the three algorithms for the simple two data sets, we were able to obtain good clustering results. Through a lot of clustering experiments for a variety of data sets, the effectiveness and reliability of the criterion and those algorithms will be investigated.

REFERENCES

[1] Duda R.O., Hart P.E., Stork D.G., *Pattern Classification* (2nd edition), John Wiley & Sons, INC., 2001.
[2] Jain A.K., Dubes R.C., *Algorithms for Clustering Data*, Prentice-Hall, Englewood Cliffs, NJ, 1988.
[3] Gordon A.D., *Classification* (2nd Edition), Chapman & Hall/CRC, 1999.
[4] Bezdek J.C., *Pattern Recognition with Fuzzy Objective Function Algorithms*, Plenum Press, NY, 1981.
[5] MacQueen J., "Some Methods for Classification and Analysis of Multivariate Observations," in Proc. 5th Berkeley Symp. on Math. Stat. and Prob. 1, Univ. of California Press, Berkeley and Los Angeles, pp. 281-297, 1967.
[6] Linde Y., Buzo A., Gray R.M., "An Algorithm for Vector Quantizer Design," IEEE Trans. Commun., vol.28, pp. 84-95, 1980.
[7] Jain A.K., "Data Clustering: 50 Years Beyond K-Means," The King-Sun Fu Prize lecture delivered at the 19th ICPR, Dec. 8, 2008.
[8] Dubes R, Jain A.K., "Validity studies in clustering methodologies," Pattern Recognition, vol. 11, no. 4, pp. 235-254, 1979.
[9] Milligan G.W., Cooper M.C., "An Examination of Procedures for Determining the Number of Clusters in a Data Set," Psychometrika, vol.50, no.2, pp. 159-179, June 1985.
[10] Bezdek J.C., Pal N.R., "Some New Indexes of Cluster Validity ," IEEE Transactions on Systems, Man, and Cybernetics, Part B:, vol.28, no. 3, pp. 301-315, Jun 1998.
[11] Calinski R. B., Harabasz J., "A Dendrite Method for Cluster Analysis," Communications in Statistics, vol. 3, no. 1, pp. 1-27, 1974.
[12] Pelleg D., Moore A., "X-means: Extending K-means with Efficient Estimation of the Number of Clusters ," in Proc. of the 17th International Conf. on Machine Learning, pp.727-734, 2000.
[13] Tibshirani R., Walther G., Hastie T., "Estimating the Number of Clusters in a Dataset via the Gap Statistic," Journal of the Royal Statistical Society: Series B, pp. 411-423, 2001.
[14] Maulik U., Bandyopadhyay S., "Performance Evaluation of Some Clustering Algorithms and Validity Indices," IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI), vol.24, no.12, pp.1650-1654, December 2002.
[15] Pakhira M.K., Bandyopadhyay S., Maulik U., "Validity Index for Crisp and Fuzzy Clusters," Pattern Recognition, vol. 37, no. 3, pp. 487-501, March 2004.
[16] Morii F., "Estimating the Number of Clusters by Distortion-Ratio Map," in Proc. of Joint 5th International Conference on Soft Computing and Intelligent Systems and 11th International Symposium on Advanced Intelligent Systems, pp. 1510-1514, 2010.
[17] Morii F., "A Clustering Criterion Based on Distortion Ratios and Its Algorithm," in Proc. of 2nd International Conference on Advanced Topics in Artificial Intelligence (ATAI 2011), pp.15-18, 2011.

Fujiki Morii received the B.E., M.E. and D.E. degrees in communiaction engineering from Osaka University in 1971, 1973 and 1976, respectively. After joining Ehime University from 1978 to 1993, he is now a Professor of the Department of Information and Computer Sciences at Nara Women's

University. He has been engaging education and research on pattern recognition, image processing and information theory. He is a member of IEEE and IEICE of Japan.