Linear Fuzzy Clustering Techniques With Missing Values and Their Application to Local Principal Component Analysis

Katsuhiro Honda, *Member, IEEE,* and Hidetomo Ichihashi, *Member, IEEE*

Abstract—In this paper, we propose two methods for partitioning an incomplete data set with missing values into several linear fuzzy clusters by extracting local principal components. One is an extension of fuzzy c-varieties clustering that can be regarded as the algorithm for the local principal component analysis of fuzzy covariance matrices. The other is a simultaneous application of fuzzy clustering and principal component analysis of fuzzy correlation matrices. Both methods estimate prototypes ignoring only missing values and they need no preprocessing of data such as the elimination of samples with missing values or the imputation of missing elements. Numerical examples show that the methods provide useful tools for interpretation of the local structures of a database.

Index Terms—Fuzzy clustering, missing value, principal component analysis.

I. INTRODUCTION

A USEFUL technique for knowledge discovery from database (KDD) is to extract local feature values by using the simultaneous applications of multivariate analysis and fuzzy clustering. Fuzzy c-varieties (FCV) proposed by Bezdek *et al.* [1]–[3] is the fuzzy clustering method that partitions a data set into several linear clusters by using linear varieties as the prototypes of the clusters. Because the FCV algorithm estimates the vectors spanning the prototypical linear varieties by solving the eigenvalue problems of the fuzzy scatter matrices, it can be said that the vectors are equivalent to local principal component vectors derived in each cluster and the algorithm performs a simultaneous application of fuzzy clustering and principal component analysis (PCA). In the algorithm, the eigenvectors corresponding to the largest eigenvalues of the fuzzy scatter matrices span the prototypical linear varieties. Using the linear combination of the objective function of FCV and that of fuzzy c-means (FCM) [1], which is useful to derive spherical clusters, the methods can be expanded to fuzzy c-elliptotypes (FCE) for the detection of ellipsoidal clusters. Though it is difficult to describe characteristics of a large-scale database by only one statistical model, we often obtain practical knowledge from a local model of each cluster.

In spite of the usefulness, however, the analyzing techniques often suffer from the presence of missing values in real world applications. For such an incomplete data set, a simple strategy

is to remove all sample data or attributes including missing values. But the strategy isn't desirable because the elimination brings a loss of information. Another technique for dealing with an incomplete data set is to impute the missing values. Besides imputing the averages of the attributes, we can yield the maximum likelihood estimates of the missing values using the EM algorithm [4] under the condition that the defects arise randomly. But the imputations are often computationally demanding and the errors in the imputation may accumulate and propagate through the iterative optimization procedures. Therefore, it is desirable to analyze the incomplete data set without elimination or imputation of data. In the application of PCA, several methods that extract principal components without any preprocessing of data have been proposed. Ruhe [5], Wiberg [6], and Shibayama [7] proposed PCA-like methods for capturing the structure of incomplete multivariate data without any imputations and statistical assumptions, and Shum *et al.* [8] applied such techniques to object modeling. The methods are based on the lower rank approximation of the data matrix, which accomplishes the minimization of the least square criterion, and they extract principal components of the covariance matrix of the sample data set. In the methods, such an iterative optimization technique as Gauss–Newton algorithm is used. Generally, PCA is not scale invariant and the results are sensitive to the scales of the measurement of the variables. Then we often use the correlation matrix instead of the covariance matrix to derive unique principal components. Takane [9] proposed a technique that can handle missing values in PCA of the correlation matrix. Takane's method can derive the solution analytically even when the data set includes missing values.

Fuzzy clustering also suffers from the presence of missing values. Miyamoto *et al.* [10] proposed several methods for handling missing values in FCM. A basic strategy is to replace missing values by the weighted averages of the corresponding attributes. Another simple approach is to ignore the missing values and calculate the distances from the remaining coordinates. Timm *et al.* [11] also proposed similar techniques and reported that the simple approach of ignoring the missing values gave fuzzier membership assignments than the strategy of replacing the missing values with the averages. Other observations such as the convergence property and the initialization problem are discussed in [12].

In this paper, we propose two methods for partitioning an incomplete data set including missing values into several linear fuzzy clusters by using local principal components. One is the direct extension of FCV to an incomplete data set and it uses the least square criterion as the objective function. FCV is the same

technique as the extraction of local principal components based on the minimization of the least square criterion, which performs the lower rank approximation of the data matrix. While the objective function of FCV is based on the minimization of the distances between data points and prototypical linear varieties, we can derive the same objective function from the least square criterion under a certain condition. The proposed objective function is equivalent to that of FCV when the data matrix includes no missing values; hence the technique is regarded as an extension of FCV into incomplete data sets.

The other is a hybrid technique of fuzzy clustering and Takane's PCA with missing values. Because the principal components are extracted from the fuzzy correlation matrix in each cluster, the method is a simultaneous application of fuzzy clustering and PCA of correlation matrices. Although the proposed objective function with no missing values is not equivalent to that of FCV, the local principal component vectors are derived by solving a set of eigenvalue problems.

In Section II, the objective function of FCV is rewritten by using the least square criterion and the FCV algorithm is extended to incomplete data sets including missing values. In Section III, a simultaneous approach to fuzzy clustering and PCA of correlation matrices, in which the objective function is defined by ignoring missing values, is proposed. Several experimental results including the knowledge discovery from point-of-sales (POS) transaction data are presented in Section IV. Section V contains the summary conclusions.

II. EXTENSION OF FCV ALGORITHM TO INCOMPLETE DATA SETS USING LEAST-SQUARE CRITERION

Let $X = (x_{ij})$ denote an $(n \times m)$ data matrix consisting of m dimensional observation of n samples where X includes missing values. In this paper, we often represent the data matrix as $X = (\boldsymbol{x}_1, \ldots, \boldsymbol{x}_m)$ using n dimensional column vectors $\boldsymbol{x}_i$'s composed of the elements of the ith columns of X, or $X = (\tilde{\boldsymbol{x}}_1, \ldots, \tilde{\boldsymbol{x}}_n)^\top$ using m dimensional column vectors $\tilde{\boldsymbol{x}}_i$'s composed of the ith row elements of X, respectively. (In the following, bold symbols represent column vectors and the column vectors superscripted by "$\sim$" are composed of the row elements of a matrix.)

Unfortunately, it is difficult to apply the existing clustering algorithms without any modification such as the completion of the data matrix. This paper deals with only the techniques that simply ignore the missing values without any elimination or imputation of data, so that the techniques are free from the influences of the preprocessing.

In this section, we use the fuzzification technique so called entropy regularization [13] instead of the weighting exponent in the standard FCM-type algorithm. Ignoring missing values, the objective function of FCM with entropy regularization and missing values is written as follows [10]:

$$L_{\text{fcmm}} = \sum_{c=1}^{C} \sum_{i=1}^{n} u_{ci} \sum_{j=1}^{m} d_{ij}(x_{ij} - b_{cj})^2$$

$$+ \lambda \sum_{c=1}^{C} \sum_{i=1}^{n} u_{ci} \log u_{ci} \quad (1)$$

where $\boldsymbol{b}_c = (b_{c1}, \ldots, b_{cm})$ is the center of the cth cluster and d_{ij} is defined by

$$d_{ij} = \begin{cases} 1, & x_{ij} \text{ is observed} \\ 0, & x_{ij} \text{ is missing.} \end{cases} \quad (2)$$

u_{ci} is the membership with the constraint

$$\sum_{c=1}^{C} u_{ci} = 1, \qquad i = 1, \ldots, n. \quad (3)$$

The larger λ is, the fuzzier the membership assignments are. The fuzzification technique has several merits, e.g., "singularities" do not occur even if several sample points are on the prototypes and cluster centers are the means of $\tilde{\boldsymbol{x}}_i$ simply weighted by u_{ci}'s. In the case of linear fuzzy clustering, the vectors spanning prototypes are also derived from standard fuzzy scatter matrices using the regularization method.

This strategy of using (1) is useful only for spherical clustering. In this paper, we enhance the method of partitioning an incomplete data set into several linear fuzzy clusters by extracting local principal components.

A. Local Principal Component Analysis Using Least Square Criterion

In this section, we show that the objective function of FCV can be rewritten by using the least square criterion introduced in PCA [14]–[16]. The goal of the simultaneous approach to PCA and fuzzy clustering is to partition the data set using local principal component vectors to express local linear structures. FCV is the clustering method that partitions a data set into C linear fuzzy clusters. The objective function of FCV with entropy regularization consists of distances from data points to p-dimensional prototypical linear varieties spanned by linearly independent vectors $\boldsymbol{a}_{cj}$'s as follows:

$$L_{fcv} = \sum_{c=1}^{C} \sum_{i=1}^{n} u_{ci} \left\{ (\tilde{\boldsymbol{x}}_i - \boldsymbol{b}_c)^\top (\tilde{\boldsymbol{x}}_i - \boldsymbol{b}_c) \right.$$

$$\left. - \sum_{j=1}^{p} \boldsymbol{a}_{cj}^\top R_{ci} \boldsymbol{a}_{cj} \right\}$$

$$+ \lambda \sum_{c=1}^{C} \sum_{i=1}^{n} u_{ci} \log u_{ci} \quad (4)$$

$$R_{ci} = (\tilde{\boldsymbol{x}}_i - \boldsymbol{b}_c)(\tilde{\boldsymbol{x}}_i - \boldsymbol{b}_c)^\top \quad (5)$$

where u_{ci} denotes the membership degree of the data point $\boldsymbol{x}_i$ to the cth cluster and $\top$ represent the transpose of the matrix or the vector.

From the necessary condition for the optimality $\partial L_{fcv}/\partial \boldsymbol{a}_{cj} = \boldsymbol{0}$, the optimal $\boldsymbol{a}_{cj}$'s are derived by solving the following eigenvalue problem:

$$S_{fc} \boldsymbol{a}_{cj} = \mu_{cj} \boldsymbol{a}_{cj} \quad (6)$$

where S_{fc} is the fuzzy scatter matrix:

$$S_{fc} = \sum_{i=1}^{n} u_{ci} R_{ci}. \quad (7)$$

Because the optimal $\boldsymbol{a}_{cj}$'s are the eigenvectors corresponding to the largest eigenvalues, the vectors are regarded as the fuzzy

principal component vectors extracted in each cluster considering the memberships [17].

In the same way, cluster centers and memberships are updated from the conditions $\partial L_{fcv}/\partial \boldsymbol{b}_c = \boldsymbol{0}$ and $\partial L_{fcv}/\partial u_{ci} = 0$ respectively. An iterative algorithm is used to derive the clustering result.

In this section, we extract local principal components by using a least square criterion. We define the least square criterion for local PCA using membership u_{ci} and entropy regularization as

$$L_{lsc} = \sum_{c=1}^{C} \text{tr}\Big\{ (X - Y_c)^\top U_c (X - Y_c) \Big\}$$
$$+ \lambda \sum_{c=1}^{C} \sum_{i=1}^{n} u_{ci} \log u_{ci} \quad (8)$$

where $U_c = \text{diag}(u_{c1}, \ldots, u_{cn})$. $Y_c = (y_{cij})$ denotes the lower rank approximation of the data matrix X in cth cluster,

$$Y_c = F_c A_c^\top + \boldsymbol{1}_n \boldsymbol{b}_c^\top \quad (9)$$

where $F_c = (\tilde{\boldsymbol{f}}_{c1}, \ldots, \tilde{\boldsymbol{f}}_{cn})^\top$ is the $(n \times p)$ score matrix and $A_c = (\boldsymbol{a}_{c1}, \ldots, \boldsymbol{a}_{cp})$ is the $(m \times p)$ principal component matrix. $\boldsymbol{1}_n$ is n dimensional vector whose all elements are 1.

With fixed memberships, the extraction of local principal components in each cluster is equivalent to the calculation of F_c, A_c and $\boldsymbol{b}_c$ such that the least square criterion of (8) is minimized.

From the necessary condition for the optimality of the objective function $\partial L_{lsc}/\partial \boldsymbol{b}_c = \boldsymbol{0}$, we have

$$\boldsymbol{b}_c = \left(\boldsymbol{1}_n^\top U_c \boldsymbol{1}_n \right)^{-1} \left(X^\top - A_c F_c^\top \right) U_c \boldsymbol{1}_n \quad (10)$$

and if $F_c^\top U_c \boldsymbol{1}_n = \boldsymbol{0}$

$$\boldsymbol{b}_c = \left(\boldsymbol{1}_n^\top U_c \boldsymbol{1}_n \right)^{-1} X^\top U_c \boldsymbol{1}_n. \quad (11)$$

Here, (11) is equivalent to the updating rule for the cluster center $\boldsymbol{b}_c$ in the FCV algorithm. Substituting (9), (8) is

$$L_{lsc} = \sum_{c=1}^{C} \Big\{ \text{tr}\left(X_c^\top U_c X_c \right) - 2\text{tr}\left(X_c^\top U_c F_c A_c^\top \right)$$
$$+ \text{tr}\left(A_c F_c^\top U_c F_c A_c^\top \right) \Big\}$$
$$+ \lambda \sum_{c=1}^{C} \sum_{i=1}^{n} u_{ci} \log u_{ci} \quad (12)$$

where $X_c = X - \boldsymbol{1}_n \boldsymbol{b}_c^\top$.

From $\partial L_{lsc}/\partial F_c = O$

$$F_c A_c^\top A_c = X_c A_c. \quad (13)$$

Under the condition that $A_c^\top A_c = I$, we have $F_c = X_c A_c$ and the objective function is transformed as follows:

$$L_{lsc} = \sum_{c=1}^{C} \Big\{ \text{tr}\left(X_c^\top U_c X_c \right) - \text{tr}\left(A_c^\top X_c^\top U_c X_c A_c \right) \Big\}$$
$$+ \lambda \sum_{c=1}^{C} \sum_{i=1}^{n} u_{ci} \log u_{ci}$$
$$= L_{fcv}. \quad (14)$$

Therefore, it can be said that (8) is equivalent to the objective function of FCV and the minimization problem is solved by computing the p largest singular values of the fuzzy scatter matrix and their associated vectors, when the data matrix does not include any missing value.

B. Extraction of Local Principal Components From Incomplete Data Sets

When we deal with an incomplete real world data set including missing values, we cannot define the objective function of FCV composed of the distances between data points and prototypical linear varieties. In this section, we propose a new clustering method that partitions an incomplete data set into several linear or ellipsoidal fuzzy clusters using the least square criterion. Because the least square criterion with a complete data set is equivalent to the objective function of FCV, the new method is an extension of FCV to incomplete data sets.

To extract principal components from an incomplete data set, Shibayama [7] proposed to estimate the lower rank approximation of the data matrix, $Y = FA^\top + \boldsymbol{1}_n \boldsymbol{b}^\top$, minimizing the following objective function:

$$\xi = \text{tr}(E^\top E) \quad (15)$$
$$E = D \odot (X - Y) \quad (16)$$

where D is the $(n \times m)$ matrix whose elements are defined by (2) and $\odot$ denotes the Hadamard product. In this method, only the elements of Y whose counterparts of X are observed are approximated while others are estimated depending on the effects of the observed elements.

The objective function to be minimized is defined by the convex combination of L_{fcmm} and L_{lsc} as follows:

$$L_{fcem} = \alpha L_{lsc} + (1 - \alpha) L_{fcmm}$$
$$= \sum_{c=1}^{C} \sum_{i=1}^{n} u_{ci}$$
$$\times \sum_{j=1}^{m} d_{ij} \Bigg\{ \alpha \left(x_{ij} - \sum_{k=1}^{p} f_{cik} a_{cjk} - b_{cj} \right)^2$$
$$+ (1 - \alpha)(x_{ij} - b_{cj})^2 \Bigg\}$$
$$+ \lambda \sum_{c=1}^{C} \sum_{i=1}^{n} u_{ci} \log u_{ci} \quad (17)$$

where α is a constant which defines the tradeoff between FCM and local PCA. When α is 0, (17) is equivalent to (1).

To obtain a unique solution, the objective function is minimized under the constraints that

$$F_c^\top U_c F_c = I, \qquad c = 1, \ldots, C \quad (18)$$
$$F_c^\top U_c \boldsymbol{1}_n = \boldsymbol{0}, \qquad c = 1, \ldots, C \quad (19)$$
$$\sum_{c=1}^{C} u_{ci} = 1; \quad i = 1, \ldots, n \quad (20)$$

and $A_c^\top A_c$ is orthogonal.

Here, the optimal solution cannot be derived from eigenvalue problems because the data matrix includes missing values. In this paper, we derive the solution based on alternating least squares.

To derive the optimal A_c and $\boldsymbol{b}_c$, we rewrite (17) as follows:

$$
\begin{aligned}
L_{\text{fcem}} = \sum_{c=1}^{C} \sum_{j=1}^{m} &\Big\{ \alpha(\boldsymbol{x}_j - F_c\tilde{\boldsymbol{a}}_{cj} - \mathbf{1}_n b_{cj})^\top U_c D_j \\
&\times (\boldsymbol{x}_j - F_c\tilde{\boldsymbol{a}}_{cj} - \mathbf{1}_n b_{cj}) \\
&+ (1-\alpha)(\boldsymbol{x}_j - \mathbf{1}_n b_{cj})^\top \\
&\times U_c D_j(\boldsymbol{x}_j - \mathbf{1}_n b_{cj}) \Big\} \\
&+ \lambda \sum_{c=1}^{C} \sum_{i=1}^{n} u_{ci} \log u_{ci}
\end{aligned}
\tag{21}
$$

where

$$
A_c = (\tilde{\boldsymbol{a}}_{c1}, \ldots, \tilde{\boldsymbol{a}}_{cm})^\top
$$

$$
D_j = \text{diag}(d_{1j}, \ldots, d_{nj}).
$$

From $\partial L_{\text{fcem}}/\partial \tilde{\boldsymbol{a}}_{cj} = \mathbf{0}$ and $\partial L_{\text{fcem}}/\partial b_{cj} = 0$, we have

$$
\tilde{\boldsymbol{a}}_{cj} = (F_c^\top U_c D_j F_c)^{-1} F_c^\top U_c D_j(\boldsymbol{x}_j - \mathbf{1}_n b_{cj})
\tag{22}
$$

$$
b_{cj} = (\mathbf{1}_n^\top U_c D_j \mathbf{1}_n)^{-1} \mathbf{1}_n^\top U_c D_j(\boldsymbol{x}_j - \alpha F_c\tilde{\boldsymbol{a}}_{cj}).
\tag{23}
$$

In the same way, we can derive the optimal F_c and u_{ci}. Equation (17) is equivalent to

$$
\begin{aligned}
L_{\text{fcem}} = \sum_{c=1}^{C} \sum_{i=1}^{n} u_{ci} &\Big\{ \alpha(\tilde{\boldsymbol{x}}_i - A_c\tilde{\boldsymbol{f}}_{ci} - \boldsymbol{b}_c)^\top \tilde{D}_i \\
&\times (\tilde{\boldsymbol{x}}_i - A_c\tilde{\boldsymbol{f}}_{ci} - \boldsymbol{b}_c) \\
&+ (1-\alpha)(\tilde{\boldsymbol{x}}_i - \boldsymbol{b}_c)^\top D_i(\tilde{\boldsymbol{x}}_i - \boldsymbol{b}_c) \Big\} \\
&+ \lambda \sum_{c=1}^{C} \sum_{i=1}^{n} u_{ci} \log u_{ci}
\end{aligned}
\tag{24}
$$

and $\partial L_{\text{fcem}}/\partial \tilde{\boldsymbol{f}}_{ci} = \mathbf{0}$ and $\partial L_{\text{fcem}}/\partial u_{ci} = 0$ yields (25) and (26), as shown at the bottom of the page, where

$$
\tilde{D}_i = \text{diag}(d_{i1}, \ldots, d_{im}).
\tag{27}
$$

The proposed algorithm can be written as follows.

```
FCE With Missing Values (FCEM) Algorithm
  Step 1) Initialize U_c, A_c, b_c, F_c ran-
          domly in each cluster and nor-
          malize them so that they sat-
          isfy the constraints (18)-(20)
          and A_c^T A_c is orthogonal.
  Step 2) Update A_c's using (22) and
          transform them so that each A_c^T A_c
          is orthogonal.
  Step 3) Update F_c's using (25) and nor-
          malize them so that they satisfy
          the constraints (18) and (19).
  Step 4) Update b_c's using (23).
  Step 5) Update U_c's using (26) and nor-
          malize them so that (20) holds.
  Step 6) If

              max | u_ci^NEW - u_ci^OLD |< ε
              i,c

          then stop. Otherwise, return to
          Step 3).
```

The orthogonalization of the matrices in Steps 2) and 3) is performed by such a technique as Gram–Schmidt's orthogonalization.

Here, we can rotate the score matrix F_c and the principal component matrix A_c applying the technique used in factor analysis. Assume that T is an arbitrary orthonormal matrix as

$$
T^\top T = TT^\top = I
\tag{28}
$$

and F_c^* and A_c^* are the transformed matrices

$$
F_c^* = F_c T
$$
$$
A_c^* = A_c T.
$$

F_c^* and A_c^* are also the solutions of the minimization problem because

$$
\begin{aligned}
Y_c &= F_c^* A_c^{*\top} + \mathbf{1}_n \boldsymbol{b}_c \\
&= F_c^\top TT^\top A_c^\top + \mathbf{1}_n \boldsymbol{b}_c \\
&= F_c A_c^\top + \mathbf{1}_n \boldsymbol{b}_c.
\end{aligned}
\tag{29}
$$

Thus, the principal component vectors derived by our method have flexibilities and we can give some meanings to the principal component scores through a suitable rotation in the interpretation of the results based on a priori knowledge.

$$
\tilde{\boldsymbol{f}}_{ci} = \left(A_c^\top \tilde{D}_i A_c \right)^{-1} A_c^\top \tilde{D}_i(\tilde{\boldsymbol{x}}_i - \boldsymbol{b}_c)
\tag{25}
$$

$$
u_{ci} = \exp\left\{ -\frac{\alpha(\tilde{\boldsymbol{x}}_i - A_c\tilde{\boldsymbol{f}}_{ci} - \boldsymbol{b}_c)^\top \tilde{D}_i(\tilde{\boldsymbol{x}}_i - A_c\tilde{\boldsymbol{f}}_{ci} - \boldsymbol{b}_c) + (1-\alpha)(\tilde{\boldsymbol{x}}_i - \boldsymbol{b}_c)^\top \tilde{D}_i(\tilde{\boldsymbol{x}}_i - \boldsymbol{b}_c)}{\lambda} - 1 \right\}
\tag{26}
$$

III. EXTRACTION OF LOCAL PRINCIPAL COMPONENTS OF CORRELATION MATRIX

There are two alternatives in fuzzy PCA [17]. One is to use fuzzy covariance matrices (or fuzzy scatter matrices) and FCV is included in this category. The other is to use fuzzy correlation matrices. In this section, we propose a linear clustering technique that is a simultaneous approach to fuzzy clustering and PCA of correlation matrices.

Takane [9] proposed a technique for the extraction of principal components from an incomplete data set by extending the linear equating method proposed by Shibayama *et al.* [18]. Let $D_{\tilde{\pmb{x}}_i}$ denotes an $(m \times m)$ diagonal matrix whose diagonal elements are the ith row elements of matrix X and Z_i is the transformation of $D_{\tilde{\pmb{x}}_i}$ as follows:

$$Z_i = D_{\tilde{\pmb{x}}_i} V + V_0, \qquad i = 1, \ldots, n \tag{30}$$

where $(m \times p)$ matrix V and $(m \times p)$ matrix V_0 are unknown weight matrices. The objective function of PCA for an incomplete data set with missing values is defined as follows:

$$L_{\text{pcam}} = \sum_{i=1}^{n} \text{tr}\left(\left(Z_i - \mathbf{1}_m \pmb{w}_i^\top\right)^\top \tilde{D}_i \left(Z_i - \mathbf{1}_m \pmb{w}_i^\top\right) \right) \tag{31}$$

where $\tilde{D}_i$ is defined by (27). To derive a unique solution, the objective function is minimized under the condition

$$V^\top S V = I \tag{32}$$

where S is the diagonal matrix composed of the diagonal elements of the scatter matrix. The objective function is minimized when all the elements of the jth column of Z_i are equal to the jth element of $\pmb{w}_i$, i.e., all elements are mutually equalized by the jth column vector of V, and the optimal solution is derived by PCA of the correlation matrix if the data set includes no missing values. Although the technique can be applied only to PCA of the correlation matrix, the optimal principal component vectors are derived by solving an eigenvalue problem even when the data set includes missing values.

In the following, we propose a linear fuzzy clustering technique by combining Takane's technique and FCM in the same way as FCE, in which the objective function is the linear combination of two criteria (FCM and local PCA of covariance matrices).

A. Fuzzy Principal Component Analysis of Correlation Matrix With Missing Values

In this section, we propose a technique for fuzzy PCA of correlation matrices when data include missing values. In the technique, memberships are introduced into Takane's technique. Assume that the data set was partitioned into C fuzzy clusters and the membership values are given.

To extract local principal components in cth cluster, we define the following linear model:

$$Z_{ci} = D_{\tilde{\pmb{x}}_i} V_c + V_{0c}, \qquad i = 1, \ldots, n \tag{33}$$

where Z_{ci} is an $(m \times p)$ matrix and V_c and V_{0c} are $(m \times p)$ weight matrices. The objective function is defined as

$$J_c = \sum_{i=1}^{n} u_{ci} \text{tr}\left(\left(Z_{ci} - \mathbf{1}_m \pmb{w}_{ci}^\top\right)^\top \tilde{D}_i \left(Z_{ci} - \mathbf{1}_m \pmb{w}_{ci}^\top\right) \right) \tag{34}$$

and is minimized under the condition

$$V_c^\top S_c V_c = I \tag{35}$$

where S_c is the diagonal matrix whose jth diagonal element s_{cj} is that of the modified fuzzy scatter matrix

$$s_{cj} = \sum_{i=1}^{n} d_{ij} u_{ci} (x_{ij} - b_{cj})^2. \tag{$$}$$

When V_c and V_{0c} are fixed, the optimal $\hat{\pmb{w}}_{ci}^\top$ is derived from $\partial J_c / \partial \pmb{w}_{ci} = \mathbf{0}$

$$\hat{\pmb{w}}_{ci}^\top = \frac{1}{m_i} \mathbf{1}_m^\top \tilde{D}_i Z_{ci} \tag{36}$$

$$m_i = \mathbf{1}_m^\top \tilde{D}_i \mathbf{1}_m. \tag{37}$$

Consequently, (34) is transformed into

$$J_c^*(V_c, V_{0c}) = \sum_{i=1}^{n} u_{ci} \text{tr}\left(Z_{ci}^\top C_i Z_{ci} \right) \tag{38}$$

where

$$C_i = (Q_{m/\tilde{D}i})^\top \tilde{D}_i Q_{m/\tilde{D}i} \tag{39}$$

$$Q_{m/\tilde{D}i} = I - \frac{1}{m_i} \mathbf{1}_m \mathbf{1}_m^\top \tilde{D}_i. \tag{40}$$

Substituting (33), the objective function is rewritten as

$$J_c^*\left(V_c, V_{0c}\right) = \text{tr}(V_c^\top P_{c1} V_c) + 2\text{tr}\left(V_c^\top P_{c2} V_{0c}\right) \\ + \text{tr}\left(V_{0c}^\top P_{c3} V_{0c}\right) \tag{41}$$

where P_{c1}, P_{c2}, and P_{c3} are given by

$$P_{c1} = \sum_{i=1}^{n} u_{ci} D_{\tilde{\pmb{x}}_i} C_i D_{\tilde{\pmb{x}}_i} \tag{42}$$

$$P_{c2} = \sum_{i=1}^{n} u_{ci} D_{\tilde{\pmb{x}}_i} C_i \tag{43}$$

$$P_{c3} = \sum_{i=1}^{n} u_{ci} C_i. \tag{44}$$

From $\partial J_c^* / \partial V_{0c} = O$, we have

$$P_{c3} V_{0c} = -P_{c2}^\top V_c. \tag{45}$$

We can obtain the optimal solution $\hat{V}_{0c}$ from (45). But the inverse matrix does not exist because of the depression of the rank of P_{c3} since the data set includes missing values. Therefore, $\hat{V}_{0c}$ is calculated by using P_{c3}^+ which is the Moore and Penrose generalized inverse of P_{c3}

$$\hat{V}_{0c} = -P_{c3}^+ P_{c2}^\top V_c. \tag{46}$$

Accordingly, (41) is

$$
\begin{aligned}
J_c^{**}(V_c) &= \mathrm{tr}\left(V_c^\top \left(P_{c1} - P_{c2}P_{c3}^+ P_{c2}^\top\right) V_c\right) \\
&= \mathrm{tr}(V_c^\top P_c V_c) \tag{47}
\end{aligned}
$$

where $P_c = P_{c1} - P_{c2}P_{c3}^+ P_{c2}^\top$. The constraint can be transformed into

$$
V_c^\top S_c V_c = V_c^\top \left(S_c^{1/2}\right)^\top \left(S_c^{1/2}\right) V_c = I. \tag{48}
$$

Then, (47) is

$$
J_c^{**}(\check{V}_c) = \mathrm{tr}\left(\check{V}_c^\top \left(S_c^{-1/2}\right)^\top P_c \left(S_c^{-1/2}\right) \check{V}_c\right) \tag{49}
$$

where

$$
\check{V}_c = S_c^{1/2} V_c \tag{50}
$$
$$
\check{V}_c^\top \check{V}_c = I. \tag{51}
$$

Minimizing (47) under the constraint of (51), we have the characteristic equation

$$
S_c^{-1/2} P_c S_c^{-1/2} \check{V}_c = \check{V}_c \Delta_c. \tag{52}
$$

Therefore, we obtain the principal components from the eigenvectors corresponding to the least eigenvalues of (52).

In the following, we show that the aforementioned technique is equivalent to PCA of the correlation matrix when the data matrix includes no missing values. For the illustrative purposes, let us consider the case of $p = 1$ and the goal is to estimate the weight vector $\boldsymbol{v}_c = (v_{c1}, \ldots, v_{cm})$. The linear models and the objective function J_c are defined as

$$
y_{cij} = v_{cj}x_{ij} + v_{0cj}, \qquad i = 1, \ldots, n, \ \ j = 1, \ldots, m \tag{53}
$$

$$
J_c = \sum_{i=1}^n \sum_{j=1}^m u_{ci}(y_{cij} - w_{ci})^2. \tag{54}
$$

From the necessary condition for the optimality, we have

$$
\hat{w}_{ci} = \frac{1}{m}\sum_{j=1}^m y_{cij} = \bar{y}_{ci} \tag{55}
$$

$$
\hat{\boldsymbol{v}}_{0c} = -\frac{\displaystyle\sum_{i=1}^n u_{ci}D_{\tilde{\boldsymbol{x}}_i}}{\displaystyle\sum_{i=1}^n u_{ci}}\boldsymbol{v}_c. \tag{56}
$$

Then, the objective function is transformed into

$$
J_c^{**}(\boldsymbol{v}_c) = \boldsymbol{v}_c^\top \left(\sum_{i=1}^n u_{ci}D_{\tilde{\boldsymbol{x}}_i}C_i D_{\tilde{\boldsymbol{x}}_i} \right.
$$
$$
\left. - \frac{\left(\displaystyle\sum_{i=1}^n u_{ci}D_{\tilde{\boldsymbol{x}}_i}\right) C_i \left(\displaystyle\sum_{i=1}^n u_{ci}D_{\tilde{\boldsymbol{x}}_i}\right)}{\displaystyle\sum_{i=1}^n u_{ci}}\right)\boldsymbol{v}_c. \tag{57}
$$

When the data matrix includes no missing values, $\tilde{D}_i = I$. Therefore, (40) is represented as

$$
\begin{aligned}
C_i &= \left(I - \frac{1}{m}\mathbf{1}_m\mathbf{1}_m^\top\right)^\top \left(I - \frac{1}{m}\mathbf{1}_m\mathbf{1}_m^\top\right) \\
&= I - \frac{1}{m}\mathbf{1}_m\mathbf{1}_m^\top \tag{58}
\end{aligned}
$$

and, substituting it, we have

$$
D_{\tilde{\boldsymbol{x}}_i}C_i D_{\tilde{\boldsymbol{x}}_i} = D_{\tilde{\boldsymbol{x}}_i}^2 - \frac{1}{m}\tilde{\boldsymbol{x}}_i\tilde{\boldsymbol{x}}_i^\top \tag{59}
$$

$$
\frac{\left(\displaystyle\sum_{i=1}^n u_{ci}D_{\tilde{\boldsymbol{x}}_i}\right) C_i \left(\displaystyle\sum_{i=1}^n u_{ci}D_{\tilde{\boldsymbol{x}}_i}\right)}{\displaystyle\sum_{i=1}^n u_{ci}}
$$

$$
= \sum_{i=1}^n u_{ci}D_{\boldsymbol{b}_c}^2 - \frac{1}{m}\sum_{i=1}^n u_{ci}\boldsymbol{b}_c\boldsymbol{b}_c^\top \tag{60}
$$

where $D_{\boldsymbol{b}_c}$ is the diagonal matrix composed of the elements of $\boldsymbol{b}_c$

$$
\boldsymbol{b}_c = \frac{\displaystyle\sum_{i=1}^n u_{ci}\tilde{\boldsymbol{x}}_i}{\displaystyle\sum_{i=1}^n u_{ci}}. \tag{61}
$$

Consequently, the objective function becomes

$$
J_c^{**} = \boldsymbol{v}_c^\top P_c \boldsymbol{v}_c \tag{62}
$$
$$
P_c = \sum_{i=1}^n u_{ci}\left(D_{\tilde{\boldsymbol{x}}_i}^2 - D_{\boldsymbol{b}_c}^2\right)
$$
$$
- \frac{1}{m}\sum_{i=1}^n u_{ci}(\tilde{\boldsymbol{x}}_i - \boldsymbol{b}_c)(\tilde{\boldsymbol{x}}_i - \boldsymbol{b}_c)^\top
$$
$$
= \mathrm{Diag}(S_{fc}) - \frac{1}{m}S_{fc} \tag{63}
$$

where S_{fc} is the fuzzy scatter matrix in the cth cluster. Considering $\mathrm{Diag}(S_{fc}) = S_c$

$$
S_c^{-1/2}P_c S_c^{-1/2} = I - \frac{1}{m}S_c^{-1/2}S_{fc}S_c^{-1/2} \tag{64}
$$
$$
= I - \frac{1}{m}R_{fc}. \tag{65}
$$

Here, R_{fc} is the fuzzy correlation matrix in cth cluster. Then, when $p = 1$, the eigenvalue problem of (52) is equivalent to

$$
S_c^{-1/2}P_c S_c^{-1/2}\check{\boldsymbol{v}}_c = \left(I - \frac{1}{m}R_{fc}\right)\check{\boldsymbol{v}}_c = \lambda\check{\boldsymbol{v}}_c \tag{66}
$$
$$
R_{fc}\check{\boldsymbol{v}}_c = (1 - \lambda)m\check{\boldsymbol{v}}_c \tag{67}
$$

and yields that the eigenvector corresponding to the smallest eigenvalue of $S_c^{-1/2}P_c S_c^{-1/2}$ is the principal component vector of the normalized data [17]. In the same way, we can extract two or more principal component vectors as the eigenvectors corresponding to the smallest eigenvalues of the fuzzy correlation matrix.

B. Simultaneous Application of Local PCA of Correlation Matrix and Fuzzy Clustering

Next, we combine the above technique with FCM to apply PCA of normalized data sets and fuzzy clustering simultaneously. In the following, we adopt the fuzzification method called "regularization by Kullback–Leibler (K–L) information" [19], [20] that is useful to partition a data set into several fuzzy clusters considering the proportions of the clusters. Using K–L information instead of entropy terms, the FCM algorithm with a regularized objective function is closely related to the EM algorithm with Gaussian mixture models, in which the capacities of clusters correspond to the *a priori* probabilities in the mixtured models. Because the local PCA technique proposed in Section III-A considers the normalization of data in each cluster, we should consider the normalized clustering criterion in which the capacities of the clusters are also optimized. Let the objective function of the proposed method be

$$L_{\text{lpcam}} = \sum_{c=1}^{C} \sum_{i=1}^{n} u_{ci} \sum_{j=1}^{m} d_{ij}$$

$$\times \left(\alpha \sum_{k=1}^{p} (y_{cijk} - w_{cik})^2 \right.$$

$$\left. + (1-\alpha)(x_{ij} - b_{cj})^2 \right)$$

$$+ \lambda \sum_{c=1}^{C} \sum_{i=1}^{n} u_{ci} \log \frac{u_{ci}}{\pi_c}$$

$$- \sum_{i=1}^{n} \gamma_i \left(\sum_{c=1}^{C} u_{ci} - 1 \right)$$

$$- \tau \left(\sum_{c=1}^{C} \pi_c - 1 \right) \tag{68}$$

where π_c represents the capacity (proportion) of the cth cluster and has the constraint that the sum of π_c's with respect to c is equal to 1. $(y_{cijk} - w_{cik})^2$ in the brackets corresponds to the objective function for local PCA introduced in Section III-A and $(x_{ij} - b_{cj})^2$ is the criterion for the FCM clustering. α is the trade-off coefficient between the two criteria. When $\alpha = 0$, we have spherical clusters. As α approaches 1, we have ellipsoidal or linear shape clusters. $u_{ci} \log u_{ci}/\pi_c$ represents the K–L information term and is added for the fuzzification instead of the weighting exponent or the entropy term. The minimization of the K–L information forces memberships u_{ci}'s to take the same value π_c that represents the capacity of the cluster. The larger λ is, the fuzzier the membership assignments are. When the capacity is large, the memberships become large, and *vice versa*. In this way, the memberships are fuzzified considering the capacities of the clusters unless all other π_c's except for only one are zero and all samples are crisply belonging to an identical cluster. γ_i and τ are the Lagrangian multipliers whose corresponding terms represent the constraints about the sum of memberships and that of capacities.

From the necessary condition for the optimality $\partial L_{\text{lpcam}}/\partial u_{ci} = 0$, we have

$$u_{ci} = \frac{\pi_c \exp(E_{ci})}{\sum_{l=1}^{C} \pi_l \exp(E_{li})} \tag{69}$$

where

$$E_{li} = -\frac{1}{\lambda} \sum_{j=1}^{m} d_{ij} \left(\alpha \sum_{k=1}^{p} (y_{lijk} - w_{lik})^2 \right.$$

$$\left. + (1-\alpha)(x_{ij} - b_{lj})^2 \right). \tag{70}$$

In the same way, the conditions $\partial L_{\text{lpcam}}/\partial \pi_c = 0$ and $\partial L_{\text{lpcam}}/\partial b_{cj} = 0$ yield

$$\pi_c = \frac{1}{n} \sum_{i=1}^{n} u_{ci} \tag{71}$$

$$b_{cj} = \frac{\sum_{i=1}^{n} u_{ci} d_{ij} x_{ij}}{\sum_{i=1}^{n} u_{ci} d_{ij}}. \tag{72}$$

The proposed algorithm can be written as follows.

```
Local PCA With Missing Values (LPCAM)
Algorithm
  Step 1) Initialize u_ci, i = 1,...,n ran-
          domly in each cluster and nor-
          malize them so that they satisfy
          the constraint of (20).
  Step 2) Solve the eigenvalue problem
          of (52) to derive the local prin-
          cipal component vectors, y_cij's
          and w_ci's for each cluster.
  Step 3) Update π_c's using (71).
  Step 4) Update b_c's using (72).
  Step 5) Update u_ci's using (69).
  Step 6) If

            max | u_ci^NEW - u_ci^OLD |< ε
            i,c

          then stop. Otherwise, return to
          Step 2).
```

C. Fuzzy Factor Loadings

To analyze the relation between the local principal components derived in each cluster and the original attributes, we can use fuzzy factor loading η_{ckj} [17] that is equivalent to the correlation coefficient between the kth principal component $f_{c \cdot k}$ and the jth attribute $x_{\cdot j}$.

$$\eta_{ckj} = \frac{Cov_c\{f_{c \cdot k}, x_{\cdot j}\}}{\sqrt{V_c\{f_{c \cdot k}\} V_c\{x_{\cdot j}\}}} \tag{73}$$

where $Cov_c\{f_{c\cdot k}, x_{\cdot j}\}$, $V_c\{f_{c\cdot k}\}$ and $V_c\{x_{\cdot j}\}$ represent the co-variance between $f_{c\cdot k}$ and $x_{\cdot j}$, the variance of $f_{c\cdot k}$ and that of $x_{\cdot j}$, respectively.

In the case of no missing values, the principal component score vector $\tilde{\boldsymbol{f}}_{ci}$ is calculated by

$$\tilde{\boldsymbol{f}}_{ci} = V_c^{\top}(\tilde{\boldsymbol{x}}_i - \boldsymbol{b}_c). \tag{74}$$

However, that cannot be applied to the incomplete case in which the sample includes missing values. It is natural to assume that the sample data is on the most likely point. In our experiments, we calculated $\tilde{\boldsymbol{f}}_{ci}$ using the following equation:

$$\tilde{\boldsymbol{f}}_{ci} = \left(V_c^{\top}\tilde{D}_i V_c\right)^{-1} V_c^{\top}\tilde{D}_i(\tilde{\boldsymbol{x}}_i - \boldsymbol{b}_c). \tag{75}$$

The equation is equivalent to (25) when the missing values are imputed as if the sample is on the prototypical linear variety or exists on the nearest point to the prototype.

IV. NUMERICAL EXPERIMENTS

In this section, we give two examples of how the proposed methods achieve the linear clustering of incomplete data sets. We first show illustrations of the application to a simple artificial data set and then the differences of the techniques are discussed through a real world application.

A. Analysis of Artificial Data

We generated an artificial data set consisting of 24 data points distributed uniformly on two lines in three-dimensional (3-D) space. Table I lists the coordinates of the samples and Fig. 1 shows its 3-D plots. In order to test the clustering abilities of the proposed methods, we performed several experiments using the data set. Table II shows the prototypes derived in each cluster. For the comparison, the principal component vectors of the LPCAM algorithm are transformed to the original coordinate system as follows:

$$\boldsymbol{a}_c = S_c^{1/2}\check{\boldsymbol{v}}_c. \tag{76}$$

Because the FCEM algorithm is a direct extension of the FCV algorithm, the two algorithms extracted the same prototypical lines except for the magnitude of the principal component vectors in the case of no missing values. And the LPCAM algorithm also captured the similar prototypical lines because the three attributes seem to have similar variances in each clusters.

After that, we made an incomplete data matrix including 33% missing elements, withholding one attribute from every sample data randomly. The withheld elements are represented by bold letters in Table I. Even though all samples lost one-third information, both methods captured two local structures properly. In this way, the two methods have the ability to capture the local linear structures of the incomplete data set and they provide similar clustering results if all attributes have identical variance in each cluster.

B. Analysis of Point of Sales (POS) Transaction Data

In this subsection, we discuss the characteristic features of the proposed methods through the analysis of a real world data set. We applied the linear fuzzy clustering algorithms introduced
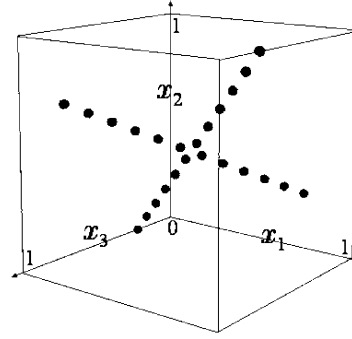

Fig. 1. 3-D plots of artificial data set.

TABLE I
COORDINATES OF ARTIFICIAL DATA SET

No.	x_1	x_2	x_3
1	**0.750**	0.250	0.000
2	0.705	**0.295**	0.091
3	0.659	0.341	**0.182**
4	0.614	**0.386**	0.273
5	**0.568**	0.432	0.364
6	0.523	0.477	**0.455**
7	**0.477**	0.523	0.545
8	0.432	0.568	**0.636**
9	0.386	0.614	**0.727**
10	0.314	**0.659**	0.818
11	**0.295**	0.705	0.909
12	0.250	**0.750**	1.000
13	**0.000**	0.000	0.250
14	0.091	0.091	**0.295**
15	0.182	0.182	**0.341**
16	**0.273**	0.273	0.386
17	0.364	**0.364**	0.432
18	**0.455**	0.455	0.477
19	0.545	**0.545**	0.523
20	0.636	**0.636**	0.568
21	0.727	**0.727**	0.614
22	0.818	0.818	**0.659**
23	**0.909**	0.919	0.705
24	1.000	1.000	**0.750**

in Sections II and III to a POS transaction data set to extract meaningful knowledge and compared the results. The POS data set collected in 1997 at two supermarkets in Osaka includes 333 sample data and each sample datum is composed of the following 20 values (meteorological elements, transaction data, and so on).

Items of POS transaction data set 1: National holiday; 2: Friday; 3: Saturday; 4: Sunday; 5: Average temperature of the day; 6, 7, 8, and 9: Temperature at 6, 12, 15, and 18 o'clock; 10: Humidity; 11 and 12: Weather category during day and night; 13: Precipitation,14, 15, and 16: Precipitation during 9–12, 12–15 and 15–18 o'clock; 17 and 18: Number of customers of supermarket A and B; 19 and 20: Sales of perishables of supermarket A and B.

The items of holiday, Friday, Saturday, and Sunday are dummy variables and weather categories have integer values (0: Sunny or Cloudy, 1: Light Rain or Shower, 2: Rain, 3: Heavy Rain). Here, "perishables" mean foods such as egg, milk, and tofu (bean curd) etc., which are made in factories and should be consumed within a few days. The goal of the analysis is to find some knowledge about the relationship among the elements.

TABLE II
RESULT OF ANALYSIS OF ARTIFICIAL DATA

algorithm	local principal component vector a_c						cluster centers b_c					
	$c=1$			$c=2$			$c=1$			$c=2$		
	x_1	x_2	x_3	x_1	x_2	x_3	x_1	x_2	x_3	x_1	x_2	x_3
FCV(0% missing)	0.666	0.667	0.334	-0.408	0.409	0.817	0.500	0.500	0.500	0.500	0.500	0.500
FCEM(0% missing)	1.085	1.086	0.544	-0.542	0.543	1.085	0.500	0.500	0.500	0.500	0.500	0.500
LPCAM(0% missing)	0.616	0.619	0.312	-0.322	0.332	0.657	0.500	0.500	0.500	0.500	0.500	0.500
FCEM(33% missing)	0.461	0.825	0.109	-0.222	0.163	0.651	0.562	0.456	0.493	0.475	0.488	0.496
LPCAM(33% missing)	0.479	0.779	0.295	-0.214	0.133	0.524	0.530	0.475	0.488	0.491	0.482	0.513

TABLE III
RESULT OF FCEM ALGORITHM WITH NO MISSING VALUES ($\alpha = 0.9$, $\lambda = 1.0$)

variable	cluster center		factor loading($c=1$)		factor loading($c=2$)	
	$c=1$	$c=2$	1st PC	2nd PC	1st PC	2nd PC
1.	-0.197	0.198	0.000	-0.003	-0.080	0.117
2.	-0.022	-0.022	0.223	**-0.383**	-0.227	**-0.449**
3.	0.000	0.000	-0.043	**0.456**	0.087	0.338
4.	0.021	0.021	-0.273	0.289	0.332	**0.428**
5.	**-0.786**	**0.790**	0.252	**0.471**	0.257	**-0.404**
6.	**-0.775**	**0.779**	0.291	**0.481**	0.255	**-0.378**
7.	**-0.765**	**0.769**	0.205	**0.468**	0.397	**-0.443**
8.	**-0.766**	**0.770**	0.112	**0.413**	0.381	**-0.472**
9.	**-0.768**	**0.771**	0.164	**0.412**	0.326	**-0.444**
10.	-0.344	0.345	**0.568**	0.262	**-0.647**	0.335
11.	-0.177	0.178	**0.687**	0.246	**-0.750**	**0.453**
12.	-0.105	0.105	**0.391**	0.034	**-0.422**	**0.421**
13.	-0.140	0.140	**0.830**	**0.449**	**-0.669**	**0.393**
14.	0.058	-0.059	**0.426**	0.040	**-0.693**	0.260
15.	0.033	-0.034	**0.698**	**0.435**	**-0.522**	0.182
16.	0.033	-0.034	**0.641**	**0.387**	-0.306	0.258
17.	-0.140	0.140	**-0.601**	**0.552**	**0.608**	**0.490**
18.	-0.162	0.163	**-0.577**	**0.687**	**0.630**	**0.608**
19.	-0.197	0.198	**-0.628**	**0.709**	**0.645**	**0.658**
20.	-0.089	0.089	**-0.554**	**0.712**	**0.562**	**0.740**

TABLE IV
RESULT OF LPCAM ALGORITHM WITH NO MISSING VALUES ($\alpha = 0.9, \lambda = 1.0$)

variable	cluster center		factor loading($c=1$)		factor loading($c=2$)	
	$c=1$	$c=2$	1st PC	2nd PC	1st PC	2nd PC
1.	-0.021	0.020	0.056	0.005	0.036	0.014
2.	0.033	-0.031	0.192	-0.163	-0.120	-0.203
3.	-0.070	0.065	-0.072	0.128	0.106	0.081
4.	-0.038	0.035	-0.120	0.144	0.077	**0.352**
5.	**-0.829**	**0.777**	**0.669**	**0.660**	**-0.841**	-0.022
6.	**-0.823**	**0.771**	**0.666**	**0.646**	**-0.822**	-0.033
7.	**-0.820**	**0.768**	**0.641**	**0.674**	**-0.854**	0.047
8.	**-0.812**	**0.761**	**0.624**	**0.678**	**-0.866**	0.065
9.	**-0.814**	**0.762**	**0.634**	**0.661**	**-0.857**	0.028
10.	-0.279	0.261	**0.488**	-0.034	0.035	**-0.497**
11.	-0.119	0.112	0.339	-0.200	0.231	**-0.544**
12.	-0.107	0.100	0.330	-0.103	0.180	-0.246
13.	-0.155	0.145	**0.488**	-0.105	0.165	**-0.525**
14.	-0.093	0.087	**0.387**	-0.094	0.113	**-0.411**
15.	-0.089	0.083	**0.430**	-0.083	0.109	**-0.408**
16.	-0.088	0.082	0.296	-0.181	0.097	-0.290
17.	-0.194	0.182	-0.195	**0.442**	-0.008	**0.588**
18.	-0.227	0.213	-0.188	**0.457**	0.025	**0.607**
19.	-0.257	0.241	-0.168	**0.536**	0.024	**0.639**
20.	-0.138	0.130	-0.252	**0.383**	0.143	**0.588**

Before the experiments, the data set was preprocessed so that each attribute had zero mean and unit variance.

First, we applied the FCEM algorithm and local PCA of the correlation matrix proposed in Section III to compare the differences between fuzzy scatter matrices and correlation matrices using whole elements of the data matrix with no missing values. Tables III and IV show the cluster center and the factor loadings of two principal components derived in each cluster. In the tables, the elements whose absolute values are greater than 0.35 are represented in bold letters.

The cluster centers in Tables III and IV show that both methods partitioned the data set into warm season and cool season by considering the temperatures(variable 5–9). From the comparison of the factor loadings, we can derive the following characteristics: the first principal components (first PCs) derived from correlation matrices are closely related to the temperatures and indicate that the temperatures have positive correlation with the precipitations only in cool season. On the other hand, the first PCs derived by the FCEM algorithm are closely related to the precipitations and the transaction data and, indicate that the numbers of customers and the sales of perishables have negative correlation with the precipitations. The difference came from the influences of the data partitioning. Because the data set was partitioned with respect to the temperatures, the variances of the temperatures in each cluster became smaller than that of other elements and the FCEM algorithm neglected the influences of the temperatures. In this way, the factor loadings of the FCEM (or FCV) algorithm can

TABLE V
RESULT OF FCEM ALGORITHM WITH MISSING VALUES ($\alpha = 0.9, \lambda = 1.0$)

variable	cluster center		factor loading($c=1$)		factor loading($c=2$)	
	$c=1$	$c=2$	1st PC	2nd PC	1st PC	2nd PC
1.	-0.018	0.021	0.042	0.032	0.001	0.020
2.	0.016	-0.026	0.164	**-0.395**	-0.327	0.114
3.	-0.040	0.052	0.002	**0.372**	0.169	0.333
4.	-0.054	0.056	0.233	0.333	**0.379**	0.270
5.	**-0.791**	**0.748**	0.300	**0.366**	-0.010	-0.176
6.	**-0.761**	**0.720**	0.302	**0.375**	-0.067	-0.166
7.	**-0.771**	**0.760**	0.190	**0.391**	0.051	-0.219
8.	**-0.775**	**0.707**	0.224	**0.369**	-0.007	-0.239
9.	**-0.797**	**0.738**	0.279	**0.359**	0.012	-0.210
10.	-0.274	0.208	**0.580**	0.195	**-0.418**	**0.452**
11.	-0.054	0.112	**0.611**	0.204	**-0.445**	**0.541**
12.	-0.107	0.099	**0.391**	0.068	-0.305	**0.397**
13.	-0.036	0.054	**0.659**	0.195	**-0.466**	**0.564**
14.	-0.067	0.104	**0.472**	0.028	**-0.391**	**0.398**
15.	0.093	-0.066	**0.505**	0.200	-0.269	0.280
16.	0.006	0.016	**0.493**	0.151	-0.269	0.280
17.	-0.206	0.229	**-0.454**	**0.611**	**0.677**	0.249
18.	-0.263	0.261	**-0.411**	**0.718**	**0.749**	**0.379**
19.	-0.263	0.261	**-0.445**	**0.739**	**0.749**	**0.366**
20.	-0.141	0.148	**-0.410**	**0.725**	**0.705**	**0.491**

be influenced by the change of the variances (the result of the partitioning) while the local PCA of correlation matrices is scale invariant.

Second, we analyzed the data set with missing values. Withholding two values from meteorological elements (variable 5–16) of every sample data randomly, we made an incomplete data set in which 10% of the elements of the data matrix are missing. Results of analysis by the FCEM algorithm and the LPCAM algorithm are shown in Tables V and VI, and

TABLE VI
RESULT OF LPCAM ALGORITHM WITH MISSING VALUES ($\alpha = 0.9$, $\lambda = 1.0$)

variable	cluster center		factor loading($c = 1$)		factor loading($c = 2$)	
	$c = 1$	$c=2$	1st PC	2nd PC	1st PC	2nd PC
1.	-0.016	0.016	0.046	0.040	0.038	-0.022
2.	0.049	-0.047	0.280	-0.073	-0.247	-0.114
3.	-0.083	0.080	-0.166	0.100	0.159	-0.001
4.	-0.062	0.059	-0.211	0.082	0.266	0.268
5.	**-0.793**	**0.742**	0.286	**0.783**	**-0.680**	**0.354**
6.	**-0.764**	**0.718**	0.295	**0.751**	**-0.664**	0.304
7.	**-0.759**	**0.755**	0.227	**0.785**	**-0.630**	**0.454**
8.	**-0.762**	**0.698**	0.269	**0.755**	**-0.647**	**0.411**
9.	**-0.791**	**0.736**	0.271	**0.769**	**-0.674**	**0.391**
10.	-0.314	0.229	**0.377**	0.163	-0.174	**-0.444**
11.	-0.106	0.138	**0.354**	0.038	-0.022	**-0.538**
12.	-0.109	0.091	0.336	0.056	0.017	-0.277
13.	-0.132	0.114	**0.421**	0.097	-0.068	**-0.504**
14.	-0.075	0.104	**0.352**	0.044	-0.073	**-0.363**
15.	-0.098	0.077	0.323	0.111	-0.059	-0.336
16.	-0.084	0.079	0.311	0.058	-0.055	-0.288
17.	-0.228	0.219	**-0.433**	0.335	0.332	**0.512**
18.	-0.263	0.252	**-0.448**	**0.359**	**0.380**	**0.505**
19.	-0.294	0.283	**-0.463**	**0.431**	**0.392**	**0.525**
20.	-0.185	0.177	**-0.479**	0.261	**0.476**	**0.421**

the elements whose absolute values are greater than 0.35 are represented in bold letters.

The cluster centers in Tables V and VI show that the methods also partitioned the incomplete data set into warm and cool seasons regardless of missing values. From the comparison of the factor loadings, we can derive the following characteristics: the FCEM algorithm extracted the principal components similar to the complete case although the absolute values of the fuzzy factor loadings of the meteorological elements are smaller than that of Table III. On the other hand, the principal components of the 1st cluster derived by the LPCAM algorithm are not so similar to those of Table IV although the principal components of the warm season have similar features. However, the principal components are more closely related to the temperatures than those of the FCEM algorithm and the LPCAM algorithm seems to be free from the influences of the data partitioning.

V. CONCLUSION

In this paper, we proposed two methods of handling missing values in the simultaneous application of principal component analysis and fuzzy clustering. One is a direct extension of the FCV algorithm in which the objective function is defined by using least square criterion and is equivalent to local PCA of fuzzy covariance matrix. The FCEM algorithm is useful to partition a data set into several linear clusters in the original coordinate system and the clustering result is derived by the FCM-like simple iterative procedure that does not include eigenvalue problems. However, the local principal components are not scale invariant. And the algorithm often yields heavy memory requirements because we must calculate the principal component scores of sample data in each iteration. If we deal with a large-scale data base, it might bring a severe problem.

The other is regarded as the simultaneous application of local PCA of fuzzy correlation matrices and fuzzy clustering, and it cannot be influenced by changes of the scales. Therefore, the LPCAM algorithm is useful to discover the knowledge from a data group without the influences of other groups. Although we can apply only to the PCA of normalized data, the LPCAM algorithm needs lower memory amount than the FCEM algorithm

because it estimates the principal component vectors from the fuzzy correlation matrix directly.

We used regularized objective functions with entropy or K–L information terms for the fuzzification of memberships. The FCEM algorithm can be also formulated by the standard approach based on a weighted exponent. However, it does not seem that the standard approach suits the LPCAM algorithm because the algorithm takes the normalization of attributes into consideration in each cluster.

The proposed methods are based on the assumption that the defects arise randomly. However, it is often the case that we cannot neglect the missing data mechanism and other approaches such as imputation or trilevel alternating optimization might work better than our methods. The future works include the introduction of such a mechanism to linear clustering techniques.

REFERENCES

[1] J. C. Bezdek, *Pattern Recognition With Fuzzy Objective Function Algorithms*. New York: Plenum Press, 1981.
[2] J. C. Bezdek, C. Coray, R. Gunderson, and J. Watson, "Detection and characterization of cluster substructure 1. Linear structure: Fuzzy *c*-lines," *SIAM J. Appl. Math.*, vol. 40, pp. 339–357, 1981.
[3] ——, "Detection and characterization of cluster substructure 2. Fuzzy *c*-varieties and convex combinations thereof," *SIAM J. Appl. Math.*, vol. 40, pp. 358–372, 1981.
[4] A. P. Dempster, N. M. Laird, and D. B. Rubin, "Maximum likelihood from incomplete data via the EM algorithm," *J. R. Stat. Soc.*, vol. B39, pp. 1–38, 1977.
[5] A. Ruhe, "Numerical computation of principal components when several observations are missing," Dept. Inform. Processing, Umea Univ., Umea, Sweden, Tech. Rep. UMINF-48-74, 1974.
[6] T. Wiberg, "Computation of principal components when data are missing," in *Proc. 2nd Symp. Computational Statistics*, 1976, pp. 229–236.
[7] T. Shibayama, "A PCA-like method for multivariate data with missing values" (in Japanese), *Jpn. J. Educat. Psychol.*, vol. 40, pp. 257–265, 1992.
[8] H. Shum, K. Ikeuchi, and R. Reddy, "Principal component analysis with missing data and its application to polyhedral object modeling," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 17, pp. 854–867, June 1995.
[9] Y. Takane, *Constrained Principal Component Analysis* (in Japanese). Tokyo, Japan: Asakura Shoten, 1995.
[10] S. Miyamoto, O. Takata, and K. Umayahara, "Handling missing values in fuzzy *c*-means," in *Proc. 3rd Asian Fuzzy Systems Symp.*, 1998, pp. 139–142.
[11] H. Timm and R. Kruse, "Fuzzy cluster analysis with missing values," in *Proc. 17th NAFIPS Int. Conf.*, 1998, pp. 242–246.
[12] R. J. Hathaway and J. C. Bezdek, "Fuzzy *c*-means clustering of incomplete data," *IEEE Trans. Syst., Man, Cybern.*, vol. 31, pp. 735–744, June 2001.
[13] S. Miyamoto and M. Mukaidono, "Fuzzy *c*-means as a regularization and maximum entropy approach," in *Proc. 7th Int. Fuzzy Systems Association World Congr.*, vol. 2, 1997, pp. 86–92.
[14] K. Pearson, "On lines and planes of closest fit to systems of points in space," *Phil. Mag.*, ser. 6, vol. 2, pp. 559–572, 1901.
[15] G. Young, "Maximum likelihood estimation and factor analysis," *Psychometrika*, vol. 6, pp. 49–53, 1940.
[16] P. Whittle, "On principal components and least square methods of factor analysis," *Skand. Akt.*, vol. 36, pp. 223–239, 1952.
[17] Y. Yabuuchi and J. Watada, "Fuzzy principal component analysis and its application," *Biomed. Fuzzy Human Sci.*, vol. 3, pp. 83–92, 1997.
[18] T. Shibayama and S. Shiba, "Application of a linear equating method to incomplete data" (in Japanese), *Jpn. J. Educat. Psychol.*, vol. 35, pp. 86–89, 1987.

[19] H. Ichihashi, K. Honda, and N. Tani, "Gaussian mixture pdf approximation and fuzzy c-means clustering with entropy regularization," in *Proc. 4th Asian Fuzzy Systems Symp.*, vol. 1, 2000, pp. 217–221.

[20] H. Ichihashi and K. Honda, "Fuzzy c-means clustering with regularization by K-L information," presented at the 10th IEEE Int. Conf. Fuzzy Systems, 2001.

Katsuhiro Honda (M'01) received the B.E., M.E., and D.Eng. degrees in industrial engineering from Osaka Prefecture University, Osaka, Japan, in 1997 and 1999, respectively.

He is currently a Research Associate in the Department of Industrial Engineering, Osaka Prefecture University. His research interests include hybrid techniques of fuzzy clustering and multivariate analysis, data mining with fuzzy data analysis, and neural networks.

Hidetomo Ichihashi (M'94) received the B.E. and D.Eng. degrees in industrial engineering from Osaka Prefecture University, Osaka, Japan, in 1971 and 1986, respectively.

From 1971 to 1981, he was with the Information System Center of Matsushita Electric Industrial Co., Ltd., Tokyo, Japan. From 1981 to 1993, he was a Research Associate, Assistant Professor, and Associate Professor at Osaka Prefecture University, where he is currently a Professor in the Department of Industrial Engineering. His fields of interest are adaptive modeling of GMDH-type neural networks, fuzzy c-means clustering with entropy regularization, data mining with fuzzy data analysis, human–machine interface, and cognitive engineering