

Fuzzy Local ICA for Extracting Independent Components Related to External Criteria

Katsuhiko Honda and Hidetomo Ichihashi

Graduate School of Engineering, Osaka Prefecture University
1-1 Gakuen-cho, Nakaku, Sakai, Osaka, 599-8531 Japan
{honda,ichi}@cs.osakafu-u.ac.jp

Abstract

Independent component analysis (ICA) is an unsupervised technique for blind source separation, and the ICA algorithms using non-gaussianity as the measure of mutual independence have been also used for projection pursuit or visualization of multivariate data for knowledge discovery in databases (KDD). However, in real applications, it is often the case that we fail to extract useful latent variables because they have no connection with predefined criterion variables. This paper proposes an enhanced technique of ICA, which extracts independent components closely related to some external criteria. Preprocessing is performed by using fuzzy regression-principal component analysis, which estimates latent variables that have high correlation with the external criteria considering local data structure.

Mathematics Subject Classification: 68T10, 62H25, 62H30, 03E72

Keywords: fuzzy clustering, independent component analysis, regression-principal component analysis

1 Introduction

Independent component analysis (ICA) [2, 11, 7] is an unsupervised technique for blind source separation, in which the goal is to reconstruct mutually independent signals. Linear ICA model uses higher order statistics than principal component analysis (PCA) to reveal the intrinsic linear structure of data sets, and the ICA algorithms using non-gaussianity as the measure of mutual independence have been also used for projection pursuit [10] or visualization of multivariate data for knowledge discovery in databases (KDD). In real applications, however, it is often the case that the goal of the analysis is to represent

the mutual relationship between the observed variables and some external criterion variables, and we fail to extract useful latent variables because they have no connection with predefined criterion variables.

When we have some external criteria, we should extract independent components considering the effects of external variables. Yanai [17] proposed PCA with external criteria that extracts latent variables uncorrelated to some external criteria, in which the influences of external criteria are first removed from a data matrix by using regression analysis. Honda and Ichihashi [4] proposed a technique for extracting independent components uncorrelated to some external criteria, in which the influences of external criteria are first removed from observed data by regression analysis before applying the ICA algorithm. Another possible approach is to extract feature values that are closely related to some external criteria. Regression-principal component analysis [13, 15] is a technique for extracting latent variables that have high correlation with the external criteria so as to keep the covariance structure between observed variables and the external criteria. So, regression-principal components are useful for representing the intrinsic feature of multivariate data sets than conventional principal components in the view point of prediction ability.

Linear ICA models are, however, often too simple for describing real-world data, and several non-linear ICA approaches that were used in conjunction with some suitable clustering algorithms have been proposed [9, 14]. Honda *et al.* [5] proposed a local ICA model that uses Fuzzy c -Varieties (FCV) clustering method [1] for extracting local independent components. The FCV algorithm partitions an observed data set into linear fuzzy clusters based on the similarities of mixing matrices, and can be regarded as local PCA [3], in which linear model estimation is performed in conjunction with fuzzy clustering. So, the FCV clustering is a simultaneous application of data partitioning and normalization of the observed variables. In [4], an extended local ICA model was proposed, in which preprocessing step was performed by a clustering algorithm for estimating local principal components that were uncorrelated to some external criteria.

This paper proposes an enhanced local ICA approach, in which the observed variables are preprocessed by using fuzzy local regression-principal component analysis [16] that is a hybrid technique of PCA, regression analysis and fuzzy clustering. The new technique is applied to knowledge discovery from POS transaction data with the goal of the analysis being to reveal the relationship between the number of customers and other elements.

The structure of this paper is as follows: Section 2 presents a brief review of ICA and introduces the fuzzy local ICA that extracts independent components considering membership degree of samples given in the preprocessing step by fuzzy clustering. In section 3, a new technique that extracts local independent components closely related to some external criteria is proposed. Several

experimental results including the knowledge discovery from POS transaction data are presented in Section 4. Section 5 contains the summary conclusions.

2 Fast ICA Algorithm and Local ICA Approaches

2.1 ICA Formulation and Fast ICA Algorithm

Denote that $\mathbf{x} = (x_1, x_2, \dots, x_M)^\top$ is an M dimensional observed data vector and $\mathbf{s} = (s_1, s_2, \dots, s_N)^\top$ is an N dimensional source signal vector corresponding to the observed data with $N \leq M$, where \top represents the transpose of vector. When the elements of source signals (s_1, s_2, \dots, s_N) are mutually statistically independent and have zero-means and unit-variances, the observed data are assumed to be the linear mixtures of s_i as $\mathbf{x} = A\mathbf{s}$, where unknown $M \times N$ matrix A is called a mixing matrix. The goal of ICA is to estimate source signals s_i , $i = 1, \dots, N$ and mixing matrix A using only the observed data \mathbf{x} .

Some ICA algorithms perform a preprocessing step of whitening and sphering before applying linear ICA. In the preprocessing step, observed data \mathbf{x} are transformed into linear combinations $\mathbf{z} = P^\top \mathbf{x}$ such that their elements z_i , $i = 1, \dots, N$ are mutually uncorrelated and all have unit variance, i.e., correlation matrix $E\{\mathbf{z}\mathbf{z}^\top\}$ is equal to unit matrix I_N . Usually, this transformation is performed by linear PCA. After the preprocessing, we have

$$\mathbf{z} = P^\top \mathbf{x} = P^\top A\mathbf{s} = W\mathbf{s},$$

where $W = P^\top A$ is an orthogonal matrix due to the assumption. Thus, the problem of finding matrix A is reduced to a simpler problem of finding an orthogonal matrix W , which gives reconstructed variables \mathbf{s} as $\mathbf{s} = W^\top \mathbf{z}$.

One measure of the mutual dependence of reconstructed variables used in ICA models is non-gaussianity. In order to measure the non-gaussianity of distribution, we can use the fourth-order cumulant or kurtosis $E\{(\mathbf{w}^\top \mathbf{z})^4\} - 3\|\mathbf{w}\|^4$. For a gaussian random variable, kurtosis is zero; for densities peaked at zero, it is positive, and for flatter densities, negative. Then the goal is to find a linear combination that has maximal or minimal kurtosis, and the objective function to be minimized or maximized is given as

$$L_{ica}(\mathbf{w}) = E\{(\mathbf{w}^\top \mathbf{z})^4\} - 3\|\mathbf{w}\|^4 + F(\|\mathbf{w}\|^2), \quad (1)$$

where $E\{\cdot\}$ denotes sample mean. The third term denotes the constraint of \mathbf{w} such that $\|\mathbf{w}\|^2 = 1$.

Hyvärinen and Oja [8] proposed the Fast ICA Algorithm that uses fixed-point iteration. The procedure is represented as follows:

Step1 Take a random initial weight vector $\mathbf{w}(0)$ of norm 1. Let $r = 1$.

Step2 Update $\mathbf{w}(r)$ using (2).

$$\mathbf{w}(r) = E\{\mathbf{z}(\mathbf{w}(r-1)^\top \mathbf{z})^3\} - 3\mathbf{w}(r-1) \quad (2)$$

Step3 Divide $\mathbf{w}(r)$ by its norm.

Step4 If $|\mathbf{w}(r)^\top \mathbf{w}(r-1)|$ is enough close to 1, stop: otherwise return to *Step2*.

Vectors $\mathbf{w}(r)$ obtained by the algorithm constitute the columns of orthogonal mixing matrix W . To estimate N independent components, we need to run this algorithm N times. We can estimate the independent components one by one by adding projection operation in the beginning of *Step3*.

2.2 Fuzzy Local ICA with FCV Clustering

Local ICA is a non-linear model, in which linear ICA models are estimated by considering local structure of data sets. Honda *et al.* [5] enhanced the Fast ICA algorithm to Fuzzy Fast ICA that can handle fuzziness in the iterative algorithm by using the FCV clustering as preprocessing. In the FCV clustering, data sets are partitioned into several linear shape clusters where each cluster is represented by prototypical linear variety of dimension $N(N < M)$ that passes through a point \mathbf{v}_c and is spanned by linearly independent unit vectors $\mathbf{p}_{c1}, \dots, \mathbf{p}_{cN}$. So, FCV-based approach is more useful for partitioning data considering differences of mixing matrices than conventional local ICA such as [9]. The objective function of FCV is composed of distances between data points and prototypical linear varieties as follows:

$$\begin{aligned} L_{fcv} = & \sum_{c=1}^C \sum_{j=1}^J u_{cj} \left\{ \|\mathbf{x}_j - \mathbf{v}_c\|^2 - \sum_{k=1}^N |\mathbf{p}_{ck}^\top (\mathbf{x}_j - \mathbf{v}_c)|^2 \right\} \\ & + \lambda \sum_{c=1}^C \sum_{j=1}^J u_{cj} \log u_{cj}, \end{aligned} \quad (3)$$

where C and J are the number of clusters and observations, respectively. u_{cj} is the degree of membership of the j th data point to the c th cluster. In [5], the memberships are fuzzified by using the entropy regularization technique [12] instead of the weighting exponent used in the standard FCV algorithm [1]. The larger the λ , the fuzzier the membership assignments. Using a three-step iterative algorithm, we can estimate the optimal fuzzy partition where each prototype corresponds to local principal subspace. So, the FCV clustering can be regarded as a technique for local PCA [3].

Before applying linear ICA in each cluster, the observed data \mathbf{x} is normalized to \mathbf{z}_c so that $E\{\mathbf{z}_c \mathbf{z}_c^\top\} = I_N$ is satisfied. Here $E\{\cdot\}$ means the

membership-weighted average. In order to perform ICA in each fuzzy cluster, the measure of non-gaussianity is also modified to fuzzy kurtosis as follows:

$$\text{fuzzy kurtosis} = \frac{\sum_{j=1}^J u_{cj} (\mathbf{w}_c^\top \mathbf{z}_{cj})^4}{\sum_{j=1}^J u_{cj}} - 3 \|\mathbf{w}_c\|^4.$$

Then, local independent components of each cluster are estimated by the Fast ICA algorithm considering memberships of observed data.

3 Extraction of Independent Components Related to External Criteria

3.1 Preprocessing by Regression-Principal Component Analysis

When we wish to derive the independent components that are relative to some external criteria, we should perform the preprocessing so that the principal components have high correlation with the external criteria. Regression-principal component analysis [13, 15] is a technique for estimating latent variables which are useful for predicting external criteria (objective variables) avoiding the problem of co-linearity.

Assume that we have J samples with I predictor variables and K response variables which are denoted by x_1, x_2, \dots, x_I and $x_{I+1}, x_{I+2}, \dots, x_{I+K}$, respectively.

$$(X | Y) = \left(\begin{array}{ccc|ccc} x_{11} & \cdots & x_{I1} & x_{(I+1)1} & \cdots & x_{(I+K)1} \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ x_{1J} & \cdots & x_{IJ} & x_{(I+1)J} & \cdots & x_{(I+K)J} \end{array} \right) (\in R^{J \times (I+K)}). \quad (4)$$

Here, X is a predictor matrix and Y is a response matrix, and each variable has zero-mean. The goal of regression-principal component analysis is to estimate the multiple regression models

$$x_{I+k} = \gamma_{k1} z_1 + \gamma_{k2} z_2 + \cdots + \gamma_{kN} z_N + e_k, \quad (5)$$

where z_1, z_2, \dots, z_N ($N < I$) are the latent variables (regression-principal components) $z_i = \mathbf{l}^\top \mathbf{x}$. The purpose is to obtain \mathbf{z} so as to minimize the sum of residual variances, i.e., we determine the coefficient vector \mathbf{l} which maximize the correlation ρ_k between the response variables x_{I+k} and linear combinations z_i of predictor variables.

$$\rho_k = (\mathbf{l}^\top \mathbf{r}_k) (\mathbf{l}^\top \mathbf{S} \mathbf{l})^{-1/2} (t_{kk})^{-1/2}, \quad (6)$$

where S is the variance-covariance matrix of the predictor variables. \mathbf{r}_k is a column of the covariance matrix R between the predictor variables and the response variables, and t_{kk} is a diagonal element of the variance-covariance matrix of the response variables.

The optimal coefficient vector \mathbf{l} is the solution of the maximization problem:

$$\max \quad \sum_{k=1}^K \rho_k m_{0k} = \mathbf{l}^\top R \mathbf{m} \quad (7)$$

$$\text{s.t.} \quad \mathbf{l}^\top S \mathbf{l} = 1 \quad (8)$$

$$\mathbf{m}^\top \mathbf{m} = 1 \quad (9)$$

where the k -th element of \mathbf{m} is given as $m_k = (t_{kk})^{-1/2} m_{0k}$, and the constraints are considered in order to derive a unique solution. Here, the optimal \mathbf{l} is derived by solving the eigenvalue problem

$$S^{-1} R R^\top \mathbf{l} = \lambda^2 \mathbf{l}. \quad (10)$$

Then, we can derive the regression-principal components that are useful for representing the intrinsic structure of the response variables, and the independent components closely related to external criteria are estimated from the latent variables by using the Fast ICA algorithm.

3.2 Local ICA with Fuzzy Clustering and Regression-Principal Component Analysis

In this subsection, an enhanced version of the local ICA model is proposed by using a simultaneous application of fuzzy clustering, PCA and multiple regression analysis [16]. In order to perform not only regression-principal component analysis but also fuzzy clustering, we consider the following C ($c = 1, \dots, C$) local multiple regression models for x_{I+1}, \dots, x_{I+K} :

$$x_{I+k} = \gamma_{ck1} z_{c1} + \gamma_{ck2} z_{c2} + \dots + \gamma_{ckN} z_{cN} + e_{ck}, \quad (11)$$

where $z_{c1}, z_{c2}, \dots, z_{cN}$ ($N < I$) are local regression-principal components:

$$z_{ci} = l_{ci1}(x_1 - v_{c1}) + \dots + l_{ciI}(x_I - v_{cI}),$$

v_{ck} denotes the center of cluster c . The goal is to estimate the coefficients l_{ck1}, \dots, l_{ckI} and fuzzy memberships u_{cj} which maximize the fuzzy correlation ρ_{ck} between the response variables and linear combinations z_{ci} of predictor variables.

$$\rho_{ck} = (\mathbf{l}_c^\top \mathbf{r}_{ck}) (\mathbf{l}_c^\top S_c \mathbf{l}_c)^{-1/2} (t_{ckk})^{-1/2}, \quad (12)$$

where

$$S_c = \{s_{cks}\}, \quad (13)$$

$$s_{cks} = \sum_{j=1}^J u_{cj}(x_{kj} - v_{ck})(x_{sj} - v_{cs}), \quad (14)$$

$$r_{cks} = \sum_{j=1}^J u_{cj}(x_{kj} - v_{ck})(x_{(I+s)j} - v_{c(I+s)}), \quad (15)$$

$$t_{cks} = \sum_{j=1}^J u_{cj}(x_{(I+k)j} - v_{c(I+k)})(x_{(I+s)j} - v_{c(I+s)}). \quad (16)$$

The objectives are to maximize the weighted sum of fuzzy correlation coefficients:

$$\sum_{c=1}^C \sum_{k=1}^K \rho_{ck} m_{c0k} = \sum_{c=1}^C \mathbf{l}_c^\top R_c \mathbf{m}_c, \quad (17)$$

where

$$m_{ck} = (t_{kk})^{-1/2} m_{c0k}, \quad (18)$$

$$R_c = \{r_{cks}\}, \quad (19)$$

and to minimize the within-group sum-of-squared-errors:

$$\sum_{c=1}^C \sum_{j=1}^J u_{cj} \|\mathbf{x}_j - \mathbf{v}_c\|^2. \quad (20)$$

To estimate unique \mathbf{l}_c and \mathbf{m}_c , the parameters are derived under the following constraints:

$$\mathbf{l}_c^\top S_c \mathbf{l}_c = 1, \quad (21)$$

$$\mathbf{m}_c^\top \mathbf{m}_c = 1, \quad (22)$$

which imply that $z_{ci} = \mathbf{l}_c^\top (\mathbf{x} - \mathbf{v}_c)$, $i = 1, \dots, N$ have unit variance and are mutually orthogonal. By the Fuzzy c -Means clustering convention [1], memberships are constrained as

$$\sum_{c=1}^C u_{cj} = 1, \quad j = 1, 2, \dots, J, c = 1, 2, \dots, C. \quad (23)$$

These objectives to be maximized and constraints can be represented by a Lagrangian function as

$$L = \sum_{c=1}^C \left[\alpha \left\{ \mathbf{l}_c^\top R_c \mathbf{m}_c - \frac{1}{2} \mu_c^l (\mathbf{l}_c^\top S_c \mathbf{l}_c - 1) - \frac{1}{2} \mu_c^m (\mathbf{m}_c^\top \mathbf{m}_c - 1) \right\} \right]$$

$$\begin{aligned}
& - (1 - \alpha) \sum_{j=1}^J u_{cj} \|\mathbf{x}_j - \mathbf{v}_c\|^2 - \lambda \sum_{j=1}^J u_{cj} \log u_{cj} \\
& - \sum_{j=1}^J \nu_j \left(\sum_{c=1}^C u_{cj} - 1 \right), \tag{24}
\end{aligned}$$

where α is a constant to define the tradeoff between regression-principal component analysis and within-group sum-of-squared-errors. u_{cj} , which takes the value from interval $[0, 1]$, is the membership of sample data j in cluster c . λ in the entropy term is a weighting parameter to specify degree of fuzziness of fuzzy clusters. μ_c^l , μ_c^m and ν_j are the Lagrangian multipliers. From the necessary condition for the optimality, the optimal v_{ci} for $i \in \{1, 2, \dots, I\}$ is given as

$$v_{ci} = \frac{\sum_{j=1}^J u_{cj} x_{ij}}{\sum_{j=1}^J u_{cj}}. \tag{25}$$

We can obtain v_{ci} for $i \in \{I+1, I+2, \dots, I+K\}$ in the same manner. From $\partial L / \partial \mathbf{l}_c = 0$ and $\partial L / \partial \mathbf{m}_c = 0$, we have

$$\mu_c = \mu_c^l = \mu_c^m = \mathbf{l}_c^\top R_c \mathbf{m}_c, \tag{26}$$

$$S_c^{-1} R_c R_c^\top \mathbf{l}_c = (\mu_c)^2 \mathbf{l}_c, \tag{27}$$

and z_c is obtained from the eigenvector \mathbf{l}_c corresponding to the maximum eigenvalue $\mu_{c,max}^2$ of the eigenvalue problem of (27).

$$z_{c1} = \mathbf{l}_{c1}^\top (\mathbf{x}^* - \mathbf{v}_c^*), \tag{28}$$

where $\mathbf{x}^* = (x_1, \dots, x_I)$ and $\mathbf{v}_c^* = (v_{c1}, \dots, v_{cI})$ and $\mathbf{l}_{c1} = \mathbf{l}_c (\mathbf{l}_c^\top S_c \mathbf{l}_c)^{-1/2}$. In the same manner, we can obtain z_{c2}, \dots, z_{cN} from the eigenvectors corresponding to the second \dots N th eigenvalues. From $\partial L / \partial u_{cj} = 0$, the memberships to clusters are obtained as

$$u_{cj} = \frac{\exp(A_{cj})}{\sum_{a=1}^C \exp(A_{aj})}, \tag{29}$$

$$\begin{aligned}
A_{aj} = & \frac{\alpha}{\lambda} \left\{ \sum_{i=1}^I \sum_{k=1}^K (x_{ij} - v_{ai})(x_{(I+k)j} - v_{a(I+k)}) l_{ai} m_{ak} \right. \\
& \left. - \mu_a \sum_{i=1}^I \sum_{k=1}^K (x_{ij} - v_{ai})(x_{kj} - v_{ak}) l_{ai} l_{ak} - \frac{1 - \alpha}{\beta} \sum_{i=1}^{I+K} (x_{ij} - v_{ai})^2 \right\}. \tag{30}
\end{aligned}$$

The algorithm is as follows [16]:

Step1 Randomly choose membership u_{cj} , $c = 1, \dots, C$, $j = 1, \dots, J$ from unit interval $[0, 1]$ so that they satisfy the probabilistic constraint of (23) and compute the cluster center vector \mathbf{v}_c by (25).

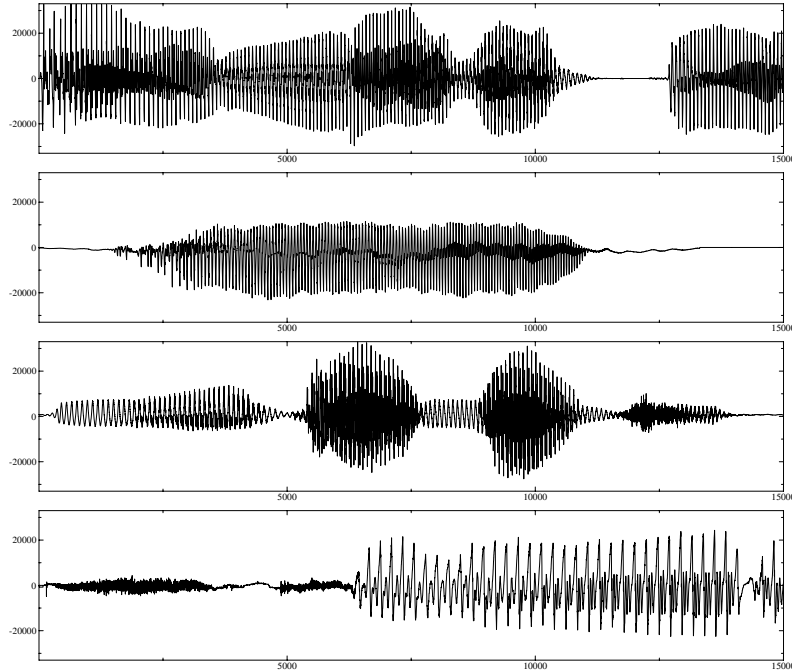


Figure 1: Original source signals

Step2 Compute $z_i, i = 1, \dots, N$ by using eigenvalues from (27).

Step3 Update \mathbf{v}_c and u_{cj} by (25) and (29).

Step4 If $\max_{c,j} |u_{cj}^{NEW} - u_{cj}^{OLD}| < \epsilon$ then stop. Otherwise, go to *Step2*.

Once we get local regression-principal components and fuzzy membership values, we can perform the Fuzzy Fast ICA algorithm in each cluster because \mathbf{z}_c is a normalized observation that satisfies the condition of $E\{\mathbf{z}_c \mathbf{z}_c^T\} = I_N$. Here, it must be noted that the proposed approach applies fuzzy linear ICA after removing components uncorrelated with the external criteria from I predictor variables while the conventional ICA model perform dimension reduction by PCA in order to remove noise components from observed data. In this sense, the independent components derived by the proposed ICA algorithm are different from conventional ones when we perform dimension reduction of observation.

4 Numerical Experiment

In this section, we show the characteristic features of the proposed method through a numerical example with an artificial data followed by a real world application to knowledge discovery from POS transaction data.

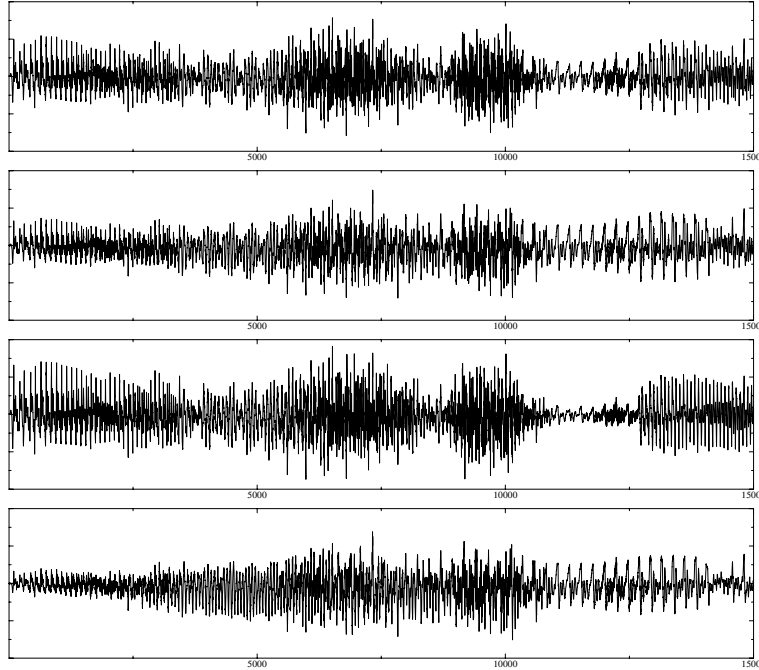


Figure 2: Mixed signals

4.1 BSS of Artificially Mixed Speech Signals

The proposed method was applied to blind source separation (BSS) of speech signals that were artificially mixed by a mixing matrix. Fig.1 and Fig.2 show the four speech signals $\mathbf{s} = (s_1, s_2, s_3, s_4)^\top$ and four mixed signals $\mathbf{x} = (x_1, x_2, x_3, x_4)^\top$ are given by $\mathbf{x} = A\mathbf{s}$, where the mixing matrix was set as

$$A = \begin{pmatrix} 0.3 & 0.2 & 0.3 & 0.2 \\ 0.2 & 0.3 & 0.2 & 0.3 \\ 0.4 & 0.2 & 0.3 & 0.1 \\ 0.1 & 0.4 & 0.2 & 0.3 \end{pmatrix}. \quad (31)$$

In this experiment, two of the original signals s_1, s_2 are given as the external criteria (response variables) x_5, x_6 with the goal of the analysis being to extract the independent components that are closely related to the external criteria, i.e., $I = 4, K = 2$. Before the application of the algorithm, the signals were normalized so that each signal has zero mean and unit variance.

First, two latent variables z_1, z_2 are extracted by regression-principal component analysis. Fig.3 shows the derived signals. Because x_3 and x_4 are not relevant to the external criteria, their influences were removed in the regression-principal component analysis step, and the regression-principal components z_1 and z_2 are the mixture of x_1 and x_2 . After that, z_1 and z_2 were transformed

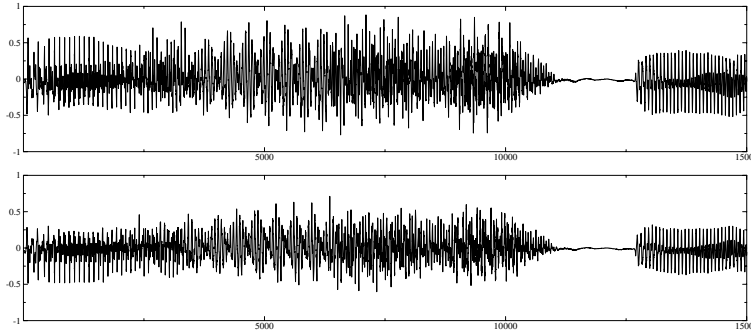


Figure 3: Regression-principal components

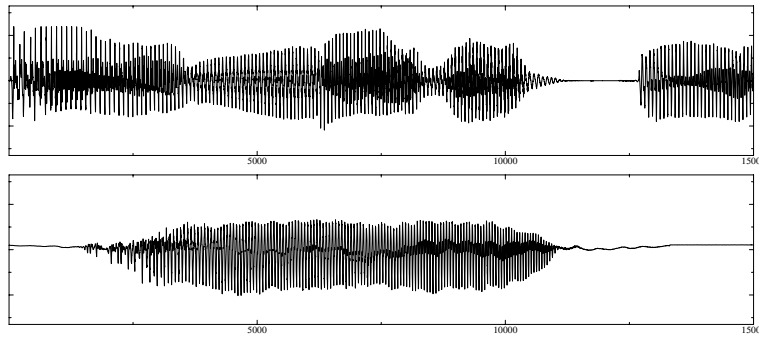


Figure 4: Reconstructed signals

into the independent components shown in Fig.4 by applying the Fast ICA algorithm. Because the regression-principal components are the latent variables that are relative to the external criteria, the reconstructed signals are very similar to the external signals of s_1, s_2 . In order to show that the proposed method is useful for extracting the independent components closely related to some external criteria, an unrealistic situation was considered, where some source signals to be reconstructed were known. In the next section, a real application is presented, where the goal is to reveal intrinsic mutual relationships from a database.

4.2 Knowledge Discovery from POS Transaction Data

Next, we applied the proposed method to knowledge discovery in a real world database. The POS (Point-Of-Sales) transaction data set, which was used in [4], was collected in 1996 at two supermarkets in Osaka and includes 327 sample data. In this experiment, we used 13 variables ($M = I + K = 13$): national holiday, 7 days of the week, average temperature of the day, humidity, precipitation and the numbers of customers in each supermarket. The items of days of the week and national holyday are dummy variables. The goal of

Table 1: Correlation coefficients between independent components and original variables derived by local ICA with FCV

Variable	Correlation coefficient			
	$c = 1$		$c = 2$	
	IC1	IC2	IC1	IC2
Holiday	0.03	-0.05	0.12	-0.36
Monday	—	—	-0.08	-0.93
Tuesday	0.83	0.54	—	—
Wednesday	-0.12	-0.64	—	—
Thursday	—	—	-0.81	0.53
Friday	0.12	-0.50	—	—
Saturday	—	—	0.89	0.40
Sunday	0.81	0.46	—	—
Temperature	0.04	0.14	-0.04	0.01
Humidity	-0.22	-0.01	-0.08	-0.30
Precipitation	-0.14	-0.20	0.21	-0.23
Supermarket A	0.17	0.82	-0.15	0.74
Supermarket B	0.34	0.87	0.54	0.50

the analysis is to extract useful knowledge on mutual relation between the numbers of customers and other elements by constructing the 2-D projections of local independent components.

For the sake of comparison, we first applied the Fuzzy Fast ICA algorithm that performs the FCV clustering as the preprocessing using all variables ($M = 13$). The parameters were set as $C = 2$, $N = 2$, $\lambda = 0.2$. In the FCV clustering stage, the data set was partitioned into two linear clusters. One mainly consisted of Tues., Wed., Fri. and Sun., and the other included the remaining days. Table 1 shows the correlation coefficients between the original variables and the independent components derived in each cluster. “—” indicates that the crisp cluster based on the maximum membership assignments did not include the day.

Fig.5 shows the projection onto the two-dimensional spaces spanned by the two local independent components (IC1 and IC2). In the figure, the horizontal and vertical axes were named based on the correlations between the independent components and the number of customers. “A-busy” (“B-busy”) means supermarket A (B) had many customers while supermarket B (A) did not have large correlation with the independent component, and vice versa. Fig.5-a shows the characteristic feature that is common to both the two supermarkets while Fig.5-b reveals the respective characteristics of each supermarket. These

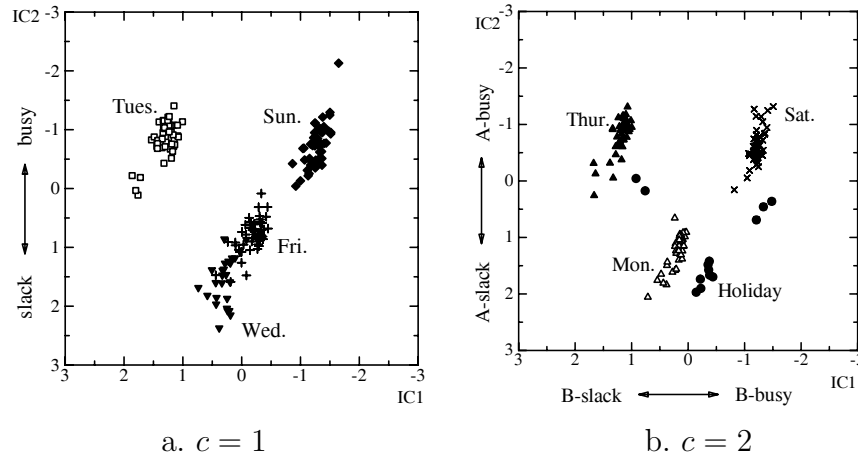


Figure 5: 2-D plots of independent components derived by local ICA with FCV

characteristic features are closely related to the average numbers of customers for each day of the week shown in Table 2. However, the independent components are not necessarily concerned with the number of customers. Indeed, IC1 of the first cluster had no information on the number of customers.

Then, we applied the proposed method in order to extract local independent components that are closely related to external criteria. The parameters were set as $C = 2$, $N = 2$, $\lambda = 0.1$, $\alpha = 0.5$. The numbers of customers were set as the external criteria ($K = 2$) and independent components were extracted from other variables ($I = 11$). Table 3 shows the correlation coefficients between the independent components and the original variables. Fig.6 shows the 2-D plots of the local independent components of each cluster. Because we used the regression-principal components that are useful for predicting the external criteria, the two independent components had high correlation with the numbers of customers in both of the two clusters. IC1 represents the characteristic feature of supermarket B while IC2 corresponds to supermarket A in the first cluster. On the other hand, in the second cluster, IC1 reveals the trend of supermarket B while IC2 shows the common feature. From the table, we can see not only the mutual relation between the numbers of customers and the day of the week but also the relation between the numbers of customers and the meteorological elements. Here, we can derive an interesting feature on the number of customers in supermarket A. Although the market had many customers on Sat. and Sun., IC1 of the first cluster had high correlation not with the days of the week but with humidity and precipitation whose averages are shown in Table 4. So, it can be said that only supermarket A was severely influenced by humidity and precipitation, and the market had many customers

Table 2: Average numbers of customers for each day of week

Day of week	Average numbers of customers	
	Supermarket A	Supermarket B
Holiday	647.9	703.7
Monday	613.4	595.9
Tuesday	693.7	671.0
Wednesday	439.4	478.3
Thursday	827.6	652.0
Friday	592.0	562.1
Saturday	748.6	750.3
Sunday	720.8	759.9

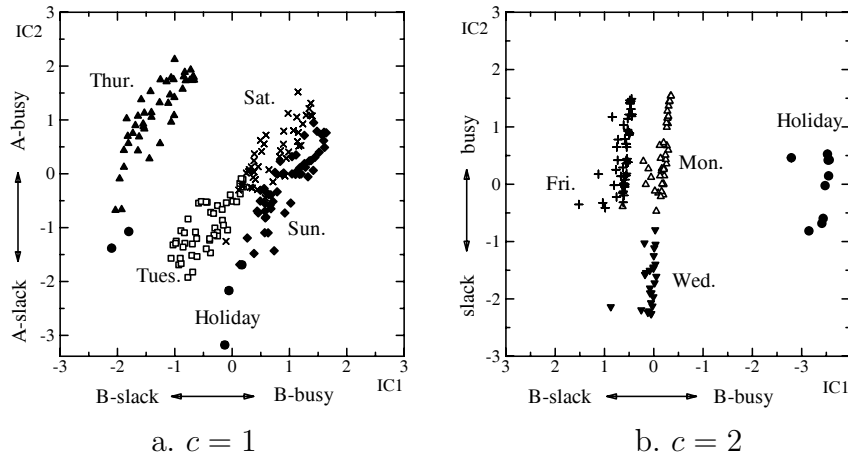


Figure 6: 2-D plots of independent components derived by proposed method

on Sat. and Sun. because we had little rain on the days as Table 4 indicates.

5 Conclusions

This paper proposed a technique for extracting independent components that are closely related to some external criteria. In the local ICA model, not only normalization of observation but also data partitioning are performed before application of linear ICA with fuzzy local regression-principal component analysis that can be regarded as a simultaneous application of fuzzy clustering, PCA and multiple regression analysis.

A potential future work is application to projection pursuit regression [6].

Table 3: Correlation coefficients between independent components and original variables derived by proposed method

Variable	Correlation coefficient			
	$c = 1$		$c = 2$	
	IC1	IC2	IC1	IC2
Holiday	-0.12	-0.31	-0.91	-0.01
Monday	—	—	-0.53	0.31
Tuesday	-0.25	-0.56	—	—
Wednesday	—	—	0.42	0.82
Thursday	-0.74	0.57	—	—
Friday	—	—	0.45	-0.39
Saturday	0.40	0.11	—	—
Sunday	0.57	-0.11	—	—
Temperature	0.34	0.34	-0.02	0.49
Humidity	0.09	-0.33	0.22	-0.09
Precipitation	-0.15	-0.31	0.25	-0.15
Supermarket A	-0.19	0.43	-0.24	0.67
Supermarket B	0.58	0.07	-0.53	0.55

Although many ICA algorithms use the non-gaussianity as the measure of mutual dependencies, the criterion is also used in projection pursuit for finding interesting distribution for visualization purposes, and the low-dimensional feature values are proved to be useful in prediction tasks. So, the proposed method can be applied to the switching projection pursuit regression tasks, as well.

References

- [1] J. C. Bezdek, *Pattern Recognition with Fuzzy Objective Function Algorithms*, Plenum Press, 1981.
- [2] P. Comon, Independent component analysis - a new concept?, *Signal Processing*, **36** (1994), 287-314.
- [3] K. Honda and H. Ichihashi, Regularized linear fuzzy clustering and probabilistic PCA mixture models, *IEEE Trans. Fuzzy Systems*, **13** (2005), 508-516.
- [4] K. Honda and H. Ichihashi, Fuzzy local independent component analysis with external criteria and its application to knowledge discovery in

Table 4: Average values of meteorological elements for each day of week

Day of week	Temp.(°C)	Hum.(%)	Prec.(mm)
Holiday	16.47	60.43	4.39
Monday	16.69	65.65	4.18
Tuesday	16.49	65.92	3.33
Wednesday	16.08	65.20	4.78
Thursday	16.84	62.38	4.74
Friday	16.30	63.22	5.59
Saturday	16.17	63.06	2.11
Sunday	16.55	63.41	2.53

databases, *International Journal of Approximate Reasoning*, **42** (2006), 159-173.

- [5] K. Honda, H. Ichihashi, M. Ohue and K. Kitaguchi, Extraction of local independent components using fuzzy clustering, *Proc. of 6th International Conference on Soft Computing* (2000), 837-842.
- [6] P. J. Huber, Projection pursuit, *The Annals of Statistics*, **13** (1985) 435-475.
- [7] A. Hyvärinen, J. Karhunen and E. Oja, *Independent Component Analysis*, John Wiley & Sons, Inc., 2001.
- [8] A. Hyvärinen and E. Oja, A fast fixed-point algorithm for independent component analysis, *Neural Computation*, **9** (1997), 1483-1492.
- [9] J. Karhunen, S. Malaroiu and M. Ilmoniemi, Local linear independent component analysis based on clustering, *Int. Jour. of Neural Systems*, **10** (2000), 439-451.
- [10] J. Karhunen, E. Oja, L. Wang, R. Vigario and J. Joutsensalo, A class of neural networks for independent component analysis, *IEEE Trans. on Neural Networks*, **8** (1997), 486-504.
- [11] T.-W. Lee, *Independent Component Analysis - Theory and Applications*, Kluwer, 1998.
- [12] S. Miyamoto and M. Mukaidono, Fuzzy *c*-Means as a regularization and maximum entropy approach, *Proc. of the 7th International Fuzzy Systems Association World Congress*, **2** (1997) 86-92.

- [13] T. Okuno, H. Kume, T. Haga and T. Yoshizawa, *Multivariable Analysis Method*, Nikka Giren, 1971. (in Japanese)
- [14] F. Palmieri and A. Budillon, Multi-class independent component analysis (MUCICA) for rank-deficient distributions, *Advances in Independent Component Analysis*, ed. by M. Girolami, Springer-Verlag, (2000), 145-160.
- [15] C. R. Rao, The use and interpretation of principal component analysis in applied research, *Sankhya, Series A*, **26** (1964), 329-358.
- [16] A. Yamakawa, K. Honda, H. Ichihashi and T. Miyoshi, Simultaneous approach to fuzzy cluster, principal component and multiple regression analysis, *Proc. of 1999 International Conference on Neural Networks* (1999), #624.
- [17] H. Yanai, Factor analysis with external criteria, *Japanese Psychol. Res.*, **12** (1970) 143-153.

Received: June 23, 2007