# Knowledge Discovery from POS Transaction Data Using Local Independent Components Uncorrelated to External Criteria

Katsuhiro Honda, Hidetomo Ichihashi
Graduate School of Engineering, Osaka Prefecture University
1-1 Gakuen-cho, Sakai, Osaka, 599-8531, Japan
honda@ie.osakafu-u.ac.jp

*Abstract*— **Local feature values derived by hybrid approaches to fuzzy clustering and multivariate data analysis have been used for knowledge discovery in databases (KDD). They often, however, fail to reveal intrinsic structure because observed variables are easily influenced by external variables. This paper proposes an enhanced technique of local independent component analysis (Local ICA), which extracts independent components uncorrelated to some external criteria. 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 days of the week.**

*Keywords*— **Fuzzy clustering, independent component analysis, external criterion, projection pursuit.**

## I. Introduction

Independent component analysis (ICA) [1], [2], [3] is an unsupervised technique, which uses higher order statistics than principal component analysis (PCA) to reveal the intrinsic structure of data sets, and is useful for projection pursuit as well [4]. Projection Pursuit has been developed in statistics for finding "interesting" features of multivariate data and its goal is to find the one-dimensional projections of multivariate data which have "interesting" distributions for visualization purposes. Typically, the interestingness is measured by the non-Gaussianity that is also used in ICA algorithms for measuring the mutual dependence of reconstructed variables. Therefore, the basis vectors of ICA should be especially useful in projection pursuit and in extracting characteristic features from natural data.

In spite of their usefulness, the linear ICA models are often too simple for describing real-world data. So, several non-linear ICA approaches that were used in conjunction with some suitable clustering algorithms have been proposed. Karhunen *et al.* proposed local ICA models [6], in which the data are grouped in several clusters based on the similarities between the observed data ahead of the preprocessing of linear ICA using some clustering algorithms such as K-means. Honda *et al.* [7] enhanced the idea to the technique that uses Fuzzy $c$-Varieties (FCV) clustering method [8] for extracting local independent components. The FCV algorithm partitions an observed data set into linear fuzzy clusters based on the similarities of mixing matrices. Because the observed data are assumed to be the linear combinations of source signals in linear ICA models, the linear clustering methods such as the FCV algorithm are suitable for the preprocessing of the ICA algorithms. The local ICA is useful not only for blind source separation (BSS) but also for knowledge discovery in databases (KDD), and have been applied to feature extraction from POS (Point-Of-Sales) transaction data [9], [10].

However, in real applications, it is often the case that we fail to reveal intrinsic structure of databases using feature values extracted by linear models because they are influenced by external variables. Yanai [11] proposed PCA with external criteria that extracts latent variables uncorrelated to some external criteria. In the technique, the influences of external criteria are first removed from a data matrix by using regression analysis. Oh *et al.* [12] enhanced the modified PCA technique for local linear modeling by using Fuzzy $c$-Regression Models (FCRM) [13] instead of regression analysis. FCRM is a switching regression model that performs fuzzy clustering and regression analysis, simultaneously.

This paper proposes an enhanced technique of local independent component analysis (Local ICA), which extracts independent components uncorrelated to some external criteria. In [7], local independent components were extracted by replacing the preprocessing of ICA with the FCV clustering. In this paper, the preprocessing is performed by using the fuzzy clustering algorithm that extracts local principal components uncorrelated to external criteria. 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 days of the week.

## II. Fast ICA Algorithm and Local ICA Approaches

### A. ICA Formulation and Fast ICA Algorithm

Denote that $\boldsymbol{y}$ is an $M$ dimensional observed data vector and $\boldsymbol{s}$ is an $N$ dimensional source signal vector corresponding to the observed data with $N \leq M$,

$$\begin{aligned}\boldsymbol{y} &= (y_1, y_2, \cdots, y_M)^\top,\\ \boldsymbol{s} &= (s_1, s_2, \cdots, s_N)^\top,\end{aligned}$$

where $\top$ represents the transpose of vector. When the elements of source signals $(s_1, s_2, \cdots, s_N)$ are mutually statistically independent and have zero-means, the observed data are assumed to be the linear mixtures of $s_i$ as follows:

$$\boldsymbol{y} = A\boldsymbol{s}, \tag{1}$$

where unknown $M \times N$ matrix $A$ is called a mixing matrix. The goal of ICA is to estimate source signals $s_i$, $i = 1, \cdots, N$ and mixing matrix $A$ using only observed data $\boldsymbol{y}$.

It is useful to apply a preprocessing of whitening and sphering by using PCA before applying the ICA algorithm [1], [14], [15]. In the preprocessing, observed data $\boldsymbol{y}$ are transformed into linear combinations $\boldsymbol{z}$

$$\boldsymbol{z} = P^\top \boldsymbol{y},$$

such that their elements $z_i$, $i = 1, \cdots, N$ are mutually uncorrelated and all have unit variance. This preprocessing implies that correlation matrix $E\{\boldsymbol{z}\boldsymbol{z}^\top\}$ is equal to unit matrix $I$, and is usually performed by PCA.

After transformation, we have

$$\boldsymbol{z} = P^\top \boldsymbol{y} = P^\top A\boldsymbol{s} = W\boldsymbol{s},$$

where $W = P^\top A$ is an orthogonal matrix due to the assumption. Thus we can reduce the problem of finding arbitrary full-rank matrix $A$ to the simpler problem of finding an orthogonal matrix $W$, which gives $\boldsymbol{s} = W^\top \boldsymbol{z}$. Hyvärinen and Oja [16] used non-Gaussianity as the measure of the mutual dependence of reconstructed variables and proposed the following objective function to be minimized or maximized.

$$L_{fica}(\boldsymbol{w}) = E\{(\boldsymbol{w}^\top \boldsymbol{z})^4\} - 3\|\boldsymbol{w}\|^4 + F(\|\boldsymbol{w}\|^2), \tag{2}$$

where $E\{(\boldsymbol{w}^\top \boldsymbol{z})^4\} - 3\|\boldsymbol{w}\|^4$ is the fourth-order cumulant or kurtosis that measures the Gaussianity of distribution. Maximizing the non-Gaussianity of reconstructed signals gives us one of independent components. The third term denotes the constraint of $\boldsymbol{w}$ such that $\|\boldsymbol{w}\|^2 = 1$.

The Fast ICA Algorithm that uses fixed-point iteration [16] is represented as follows:

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

Step2 Update $\boldsymbol{w}(r)$ using Eq.(3).

$$\boldsymbol{w}(r) = E\{\boldsymbol{z}(\boldsymbol{w}(r-1)^\top \boldsymbol{z})^3\} - 3\boldsymbol{w}(r-1). \tag{3}$$

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

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

Vectors $\boldsymbol{w}(r)$ 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.

### *B. Fuzzy Local ICA with FCV Clustering*

Honda *et al.* [7] enhanced the Fast ICA algorithm to Fuzzy Fast ICA that can handle fuzziness in the iterative algorithm by using the FCV clustering as preprocessing. The FCV clustering simultaneously performs fuzzy clustering and PCA. The principal subspace of each cluster is estimated as the prototypical linear variety of dimension $N$ that passes through point $\boldsymbol{v}_c$ and is spanned by linearly independent vectors $\boldsymbol{p}_{c1}, \cdots, \boldsymbol{p}_{cN}$. The objective function of FCV is composed of distances between data points and prototypical linear varieties as follows:

$$\begin{aligned} L_{fcv} &= \sum_{c=1}^{C}\sum_{i=1}^{J} u_{ci}\{\|\boldsymbol{x}_i - \boldsymbol{v}_c\| - \sum_{k=1}^{N} |\boldsymbol{p}_{ck}^\top(\boldsymbol{x}_i - \boldsymbol{v}_c)|^2\} \\ &\quad +\lambda \sum_{c=1}^{C}\sum_{i=1}^{J} u_{ci}\log u_{ci}, \end{aligned} \tag{4}$$

where $C$ and $J$ are the number of clusters and observation, respectively. $u_{ci}$ is the degree of membership of the $i$th data point to the $c$th cluster. In [7], the memberships are fuzzified by using the entropy regularization technique [17] instead of the weighting exponent used in the standard FCV algorithm. The larger the $\lambda$, the fuzzier the membership assignments. The fuzzification technique has several merits, e.g., "singularities" do not occur even if several sample points are on prototypes and cluster centers are the means of $\boldsymbol{x}_i$ simply weighted by $u_{ci}$'s. Using a three-step iterative algorithm, we can estimate the optimal fuzzy partition where prototypes of clusters are linear varieties. Because the optimal $\boldsymbol{p}_{ck}$ is derived as the fuzzy principal component vector of the $c$th cluster, the FCV clustering can be regarded as a technique for local PCA [18].

In order to perform ICA in each fuzzy cluster, observed data $\boldsymbol{y}$ is normalized to $\boldsymbol{z}_c$ so that $E\{\boldsymbol{z}_c\boldsymbol{z}_c^\top\} = I$ is satisfied in each cluster where $E\{\cdot\}$ means the weighted average. The measure of non-Gaussianity is also modified to fuzzy kurtosis as follows:

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

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

## III. Local Independent Component Analysis with External Criteria

### *A. Preprocessing with Principal Components Uncorrelated to External Criteria*

Assume that $M$ observed variables $y_1, \cdots, y_M$ are influenced by $K$ external criteria $x_1, \cdots, x_K$ and the two data matrices are given as follows:

$$X = \begin{pmatrix} \boldsymbol{x}_1^\top \\ \boldsymbol{x}_2^\top \\ \vdots \\ \boldsymbol{x}_J^\top \end{pmatrix} = \begin{pmatrix} x_{11} & x_{12} & \dots & x_{1K} \\ x_{21} & x_{22} & \dots & x_{2K} \\ \vdots & \vdots & \ddots & \vdots \\ x_{J1} & y_{J2} & \dots & x_{JK} \end{pmatrix},$$

$$Y = \begin{pmatrix} \boldsymbol{y}_1^\top \\ \boldsymbol{y}_2^\top \\ \vdots \\ \boldsymbol{y}_J^\top \end{pmatrix} = \begin{pmatrix} y_{11} & y_{12} & \cdots & y_{1M} \\ y_{21} & y_{22} & \cdots & y_{2M} \\ \vdots & \vdots & \ddots & \vdots \\ y_{J1} & y_{J2} & \cdots & y_{JM} \end{pmatrix},$$

where each variable is normalized as

$$\sum_{j=1}^{J} x_{jk} = 0 \quad ; \quad k = 1, \cdots, K, \tag{5}$$

$$\sum_{j=1}^{J} y_{jl} = 0 \quad ; \quad l = 1, \cdots, M. \tag{6}$$

In this subsection, we consider to extract independent components ignoring the influences of external criteria $x_1, \cdots, x_K$, i.e., we try to estimate independent components uncorrelated to the external criteria.

Yanai [11] proposed a technique for estimating principal components that are independent to some external criteria. In the technique, influences of external criteria are first removed from observed variables by considering the following linear regression model.

$$\boldsymbol{y}^\top = \boldsymbol{x}^\top B + \boldsymbol{e}^\top, \tag{7}$$

where $B$ is the partial regression coefficient matrix,

$$B = \begin{pmatrix} b_{11} & b_{12} & \ldots & b_{1M} \\ b_{21} & b_{22} & \ldots & b_{2M} \\ \vdots & \vdots & \ddots & \vdots \\ b_{K1} & b_{K2} & \ldots & b_{KM} \end{pmatrix}, \tag{8}$$

and $\boldsymbol{e}$ is the error term. Minimizing the least squares criterion

$$\begin{aligned} L_{ra}(B) &= \sum_{j=1}^{J} \|\boldsymbol{y}_j^\top - \boldsymbol{x}_j^\top B\|^2 \\ &= \|Y - XB\|^2 \\ &= \mathrm{tr}\Big((Y - XB)^\top (Y - XB)\Big), \end{aligned} \tag{9}$$

the optimal $B$ is estimated as

$$B = (X^\top X)^{-1} X^\top Y, \tag{10}$$

and data matrix $Y$ is decomposed into

$$\begin{aligned} Y &= X(X^\top X)^{-1} X^\top Y \\ &\quad + \Big(I - X(X^\top X)^{-1} X^\top\Big) Y \\ &= Y_X + Y_{\bar{X}}, \end{aligned} \tag{11}$$

where the first term of Eq.(11) is the matrix composed of the elements predictable perfectly by the external criteria and the second term is composed of the elements independent of $X$. So covariance matrix $C_{YY}$ is also decomposed into

$$\begin{aligned} C_{YY} &= \frac{1}{J} Y^\top Y \\ &= \frac{1}{J} Y^\top (Y_X + Y_{\bar{X}}) \\ &= \frac{1}{J} Y^\top X (X^\top X)^{-1} X^\top Y \\ &\quad + \frac{1}{J}\Big(Y^\top Y - Y^\top X (X^\top X)^{-1} X^\top Y\Big) \\ &= C_{YX} C_{XX}^{-1} C_{XY} \\ &\quad + (C_{YY} - C_{YX} C_{XX}^{-1} C_{XY}) \\ &= C_{YY.X} + C_{YY.\bar{X}}, \end{aligned} \tag{12}$$

$$C_{YY.X} = C_{YX} C_{XX}^{-1} C_{XY}, \tag{13}$$

$$C_{YY.\bar{X}} = C_{YY} - C_{YX} C_{XX}^{-1} C_{XY}, \tag{14}$$

where $C_{XX}$ is the variance covariance matrix of $X$, and $C_{XY}$ is the covariance matrix of $X$ and $Y$. Here, it can be said that we can estimate principal components of $C_{YY}$ by analyzing $C_{YY.X}$ and $C_{YY.\bar{X}}$ separately and the factors extracted from $C_{YY.\bar{X}}$ is free from the influences of external criteria [11].

In this paper, we consider to apply the ICA algorithm to principal components extracted by Yanai's technique. Let $P = (\boldsymbol{p}_1, \boldsymbol{p}_2, \cdots, \boldsymbol{p}_N)$ be the $N \times N$ matrix composed of $N$ principal eigenvectors of $C_{YY.\bar{X}}$. Normalized data $Z_{\bar{X}}$ to be analyzed is calculated as

$$\begin{aligned} Z_{\bar{X}} &= Y_{\bar{X}} P \\ &= \Big(I - X(X^\top X)^{-1} X^\top\Big) Y P. \end{aligned} \tag{15}$$

Here,

$$C_{XZ_{\bar{X}}} = \frac{1}{J} X^\top \Big(I - X(X^\top X)^{-1} X^\top\Big) Y P = O$$

implies that the covariance matrix of $Z_{\bar{X}}$ and $X$ is $O$ (zero matrix), i.e., the normalized data are uncorrelated to the external criteria.

If we apply the ICA algorithm to this normalized data by multiplying orthogonal matrix $W$, reconstructed data matrix $S$ are derived as

$$S = Z_{\bar{X}} W, \tag{16}$$

and the covariance matrix of $S$ and $X$ is calculated as

$$C_{XS} = C_{XZ_{\bar{X}}} W = O. \tag{17}$$

Therefore, the independent components are also uncorrelated to the external criteria. In this way, we can extract independent components uncorrelated to some external criteria by removing the influences of the external criteria in the preprocessing with PCA.

### B. *Extraction of Local Independent Components Uncorrelated to External Criteria*

In this subsection, we consider generalization of the ICA technique proposed in the previous subsection to local ICA. Oh *et al.* [12] proposed a fuzzy clustering algorithm that extracts local principal components uncorrelated to external criteria. In the clustering technique, Fuzzy $c$-Regression Models (FCRM) [13], which simultaneously performs fuzzy clustering and regression analysis, is used not only for partitioning a data set into fuzzy clusters but also for removing the influences of external criteria from fuzzy scatter matrices. Solving the eigenvalue problems of the fuzzy scatter matrices, local principal components uncorrelated to external criteria are derived. In the following, we try to estimate local independent components from the local principal components.

FCRM is a switching regression model that estimates local linear models. In order to use the switching regression model as the preprocessing of PCA, Oh *et al.* modified the local linear models as

$$y_{cjl} = \sum_{k=1}^{K} x_{cjk} b_{ckl} + e_{cjl}, \quad (18)$$

where $x_{cjk} = x_{jk} - v^x_{ck}$ and $y_{cjl} = y_{jl} - v^y_{cl}$. $\boldsymbol{v}^x_c = (v^x_{c1}, \cdots, v^x_{cK})^\top$ and $\boldsymbol{v}^y_c = (v^y_{c1}, \cdots, v^y_{cM})^\top$ are the center of the $c$th cluster. The objective function of FCRM is defined as

$$\begin{aligned} L_{fcrm} &= \sum_{c=1}^{C}\sum_{j=1}^{J} u_{cj} \sum_{l=1}^{M} (y_{cjl} - \sum_{k=1}^{K} x_{cjk} b_{ckl})^2 \\ &\quad + \lambda \sum_{c=1}^{C}\sum_{j=1}^{J} u_{cj} \log u_{cj} \\ &= \sum_{c=1}^{C} \mathrm{tr}\Big( (Y_c - X_c B_c)^\top D_c (Y_c - X_c B_c) \Big) \\ &\quad + \lambda \sum_{c=1}^{C}\sum_{j=1}^{J} u_{cj} \log u_{cj}, \end{aligned} \quad (19)$$

where $Y_c = \{y_{cjl}\}$ and $X_c = \{x_{cjk}\}$. $D_c$ is the diagonal matrix whose the $j$th diagonal element is $u_{cj}$. $B_c$ is the partial regression coefficient matrix of the $c$th cluster. Necessary condition for the optimality $\partial L_{fcrm}/\partial B_c = O$ reduces the optimal $B_c$ as

$$B_c = (X_c^\top D_c X_c)^{-1} X_c^\top D_c Y_c. \quad (20)$$

In the same way, the optimal $u_{ck}$, $\boldsymbol{v}^x_c$ and $\boldsymbol{v}^y_c$ are derived from $\partial L_{fcrm}/\partial u_{cj} = 0$, $\partial L_{fcrm}/\partial \boldsymbol{v}^x_c = \mathbf{0}$ and $\partial L_{fcrm}/\partial \boldsymbol{v}^y_c = \mathbf{0}$, respectively. A three-step iterative algorithm derives the optimal data partition and local linear models.

Using the local linear models, the data matrix to be analyzed is decomposed in each cluster as follows:

$$\begin{aligned} Y_c &= X_c B_c + (Y_c - X_c B_c) \\ &= Y_{X_c} + Y_{\bar{X}_c}. \end{aligned} \quad (21)$$

Then, the fuzzy scatter matrix of the $c$th cluster is also decomposed as

$$\begin{aligned} S_{fc} &= Y_c^\top D_c Y_c \\ &= Y_c^\top D_c (Y_{X_c} + Y_{\bar{X}_c}) \\ &= Y_c^\top D_c X_c B_c + \{Y_c^\top D_c Y_c - Y_c^\top D_c X_c B_c\} \\ &= S_{fcX} + S_{fc\bar{X}}, \end{aligned} \quad (22)$$

where

$$S_{fcX} = Y_c^\top D_c X_c (X_c^\top D_c X_c)^{-1} X_c^\top D_c Y_c, \quad (23)$$

$$\begin{aligned} S_{fc\bar{X}} &= Y_c^\top D_c Y_c \\ &\quad - Y_c^\top D_c X_c (X_c^\top D_c X_c)^{-1} X_c^\top D_c Y_c. \end{aligned} \quad (24)$$

Here, $S_{fcX}$ is closely related to $X_c$ while $S_{fc\bar{X}}$ is independent to $X_c$. So, the local principal components uncorrelated to the external criteria are derived by solving the eigenvalue problems of $S_{fc\bar{X}}$, and local independent components are estimated from the local principal components by using the Fuzzy Fast ICA algorithm.

The procedure for extracting local independent components uncorrelated to external criteria can be written as follows:

Step1 Perform the FCRM clustering to estimate local regression models.

Step2 Perform the whitening of observed data in each cluster using local PCA with external criteria.

Step3 Extract local independent components of each cluster using Fuzzy Fast ICA algorithm.

## IV. Knowledge Discovery from POS Transaction Data

In this section, we show the characteristic features of the proposed method through a real world application to knowledge discovery from POS (Point-Of-Sales) transaction data. The POS transaction data set, which was used in [10], was collected in 1997 at two supermarkets in Osaka and includes 333 sample data. Each sample datum is composed of 10 values: national holiday, 7 days of the week and the numbers of customers in each supermarket. The items of days of week and national holyday are dummy variables. The goal of the analysis is to extract useful knowledge from the 2-D projections of local independent components.

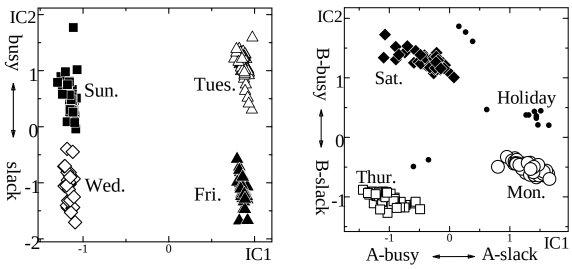
First, we briefly review the result shown in [10]. The ICA algorithms are useful for projection pursuit, in which one tries to describe the structure of high dimensional data by projecting them onto a low dimensional subspace. The structure of the POS transaction data set, however, was too complicated to derive effective 2-D projection with single linear ICA model. Then, the Fuzzy Fast ICA algorithm [7] was applied to interpret the local structure of the data set intuitively following the preprocessing with the FCV clustering. In the FCV clustering stage, the data set was partitioned into two linear clusters. One mainly consisted of Tuesday, Wednesday, Friday and Sunday, and the other included the remaining days. Table I shows the correlation coefficients of independent components and original variables derived in each cluster. In the table, "—" indicates that the cluster did not include the day when we partitioned the data based on the maximum membership assignments.

Fig.1 shows the projection onto the two-dimensional space 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

TABLE I

Correlation coefficients of independent components and original variables derived by local ICA with FCV

| Variable | Correlation coefficient | | | |
|---|---|---|---|---|
| | $c=1$ | | $c=2$ | |
| | IC1 | IC2 | IC1 | IC2 |
| Holiday | — | — | 0.27 | 0.17 |
| Monday | — | — | 0.93 | -0.26 |
| Tuesday | 0.55 | 0.69 | — | — |
| Wednesday | -0.47 | -0.43 | — | — |
| Thursday | — | — | -0.69 | -0.69 |
| Friday | 0.54 | -0.07 | — | — |
| Saturday | — | — | -0.25 | 0.93 |
| Sunday | -0.73 | 0.33 | — | — |
| Supermarket A | 0.06 | 0.70 | -0.75 | -0.04 |
| Supermarket B | -0.21 | 0.76 | -0.41 | 0.62 |




a. $c=1$ b. $c=2$

Fig. 1. Projection of independent components derived by local ICA with FCV

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. For example, in the second cluster, IC1 has the negative correlation with the number of customers of supermarket A while IC2 has positive correlation with the number of customers of supermarket B. Fig.1-a shows the characteristic feature that is common to both the two supermarkets. The number of customers is large on Sunday and Tuesday but small on Wednesday and Friday. On the other hand, Fig.1-b reveals the respective characteristics of each supermarket. The number of customers is large at supermarket A but small at supermarket B on Thursday and the supermarket B is especially busy on holiday because •'s are located above each mass in the second clusters. These characteristic features are closely related to the average numbers of customers for each day of the week shown in Table II. In this way, the local linear approach is useful for revealing the characteristic features of large scale databases.

However, it seems that they are influenced by other elements since the data of each day of the week form a long and slender mass in Fig.1. In this experiment, we tried to reveal intrinsic relationship between day of the week and the number of customers by removing influences of external criteria. We picked up three meteorological elements (Average temperature of the day, Humidity, Precipitation) as the external criteria, and partitioned the data set into two clusters using local PCA with external criteria [12]. The dimensionality of principal subspace and the coefficient of entropy term were set to be 2 and 2.0, respectively. One cluster mainly consisted of Tuesday, Thursday, Saturday and Sunday, and the other included the remaining days. We applied the Fuzzy Fast ICA algorithm to the local principal components uncorrelated to the external criteria in each cluster. Table III shows the correlation coefficients of the independent components and the original variables. Fig.2 shows the 2-D plots of the local independent components of each cluster. In Fig.2, the data of each day of the week form a spherical mass. It is because the intrinsic relationships between the variables are emphasized by removing the influences of the external criteria. Fig.2-a indicates that the number of customers is large on Saturday and Sunday but small on Thursday only in the supermarket B.

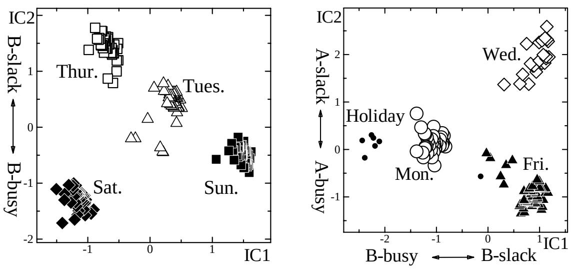
TABLE II

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 |

These intrinsic features are different from the knowledge derived in [10]. Especially, in supermarket A, the knowledge on Friday, Saturday and Sunday is in conflict with the previously extracted ones. The differences can be interpreted by taking account of the average values of the meteorological elements for each day of the week as shown in Table IV. The table indicates that Friday had many rainy days and it was few on Saturday and Sunday. Then, it can be said that the numbers of customers in supermarket A were greatly influenced by the meteorological elements and the number of customers must have been large on Fridays if we had little rain on Fridays. On the other hand, the numbers of customers in supermarket B were only slightly influenced by the external elements.

As we have demonstrated above, the proposed local ICA technique is useful for extracting the intrinsic knowledge from large scale databases by using it together with conventional projection pursuing methods.




a. $c = 1$ b. $c = 2$

Fig. 2. Projection of independent components derived by local ICA with FCRM

TABLE III

Correlation coefficients of independent components and original variables derived by local ICA with FCRM

| Variable | Correlation coefficient | | | |
|---|---|---|---|---|
| | $c = 1$ | | $c = 2$ | |
| | IC1 | IC2 | IC1 | IC2 |
| Holiday | — | — | -0.47 | 0.01 |
| Monday | — | — | -0.92 | 0.08 |
| Tuesday | 0.20 | 0.28 | — | — |
| Wednesday | — | — | 0.42 | 0.86 |
| Thursday | -0.38 | 0.77 | — | — |
| Friday | — | — | 0.62 | -0.76 |
| Saturday | -0.66 | -0.73 | — | — |
| Sunday | 0.83 | -0.28 | — | — |
| Supermarket A | -0.27 | 0.22 | -0.39 | -0.56 |
| Supermarket B | 0.14 | -0.51 | -0.53 | -0.41 |

## V. Conclusion

In this paper, we proposed a technique for extracting independent components uncorrelated to some external criteria. In the technique, the influences of the external criteria are first removed from observed data by regression analysis before applying the ICA algorithm. In the BSS problem, the preprocessing can be regarded as removing the noise signals that are given in advance.

The proposed technique was also extended to local ICA by using the FCRM clustering that is a switching regression method. In a real world application, we demonstrated that intrinsic knowledge can be discovered from large scale databases by using the proposed local ICA technique together with the conventional projection pursuing methods.

Although independent components uncorrelated to external criteria were extracted in this paper, we can also estimate independent components that are closely related to the external criteria using the remaining parts of fuzzy scatter matrices. Potential future works may include projection pursuit regression [5].

## References


[1] P. Comon, Independent component analysis - A new concept?, *Signal Processing*, vol. 36, pp. 287–314, 1994.

[2] T.-W. Lee, *Independent Component Analysis - Theory and Applications*, Kluwer, 1998.


TABLE IV

Average values of meteorological elements for each day of week

| Day of week | Temp.(℃) | 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 |


[3] A. Hyvärinen, J. Karhunen and E. Oja, *Independent Component Analysis*, John Wiley & Sons, Ing., 2001.

[4] 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*, vol. 8, pp. 486–504, 1997.

[5] P. J. Huber, Projection pursuit, *The Annals of Statistics*, vol. 13, no. 2, pp. 435–475, 1985.

[6] J. Karhunen, S. Malaroiu and M. Ilmoniemi, Local linear independent component analysis based on clustering, *Int. Jour. of Neural Systems*, vol. 10, no. 6, pp. 439–451, 1999.

[7] 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 (IIZUKA2000)*, pp. 837-842, 2000.

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

[9] K. Honda and H. Ichihashi, Feature extraction from POS transaction data by using local independent components, *Knowledge-Based Intelligent Information Engineering Systems & Allied Technologies, ed. by N. Baba, L. C. Jain and R. J. Howlett*, IOS Press, pp. 27–31, 2001.

[10] K. Honda: Studies on Local Subspace Learning for Missing Value Estimation and Exploratory Data Analysis; Ph.D. dissertation, Osaka Pref. Univ., 2004.

[11] H. Yanai, Factor analysis with external criteria, *Japanese Psychol. Res.*, vol. 12, no. 4, pp. 143–153, 1970.

[12] C.-H. Oh, K. Honda and H. Ichihashi, Knowledge discovery from local principal components independent of arbitrary factors, *Knowledge-Based Intelligent Information Engineering Systems & Allied Technologies, ed. by N. Baba, L. C. Jain and R. J. Howlett*, IOS Press, pp. 17-21, 2001.

[13] R. J. Hathaway and J. C. Bezdek, Switching regression models and fuzzy clustering, *IEEE Trans. on Fuzzy Systems*, vol. 1, no. 3, pp. 195–204, 1993.

[14] J.-F. Cardoso, Eigen-structure of the fourth-order cumulant tensor with application to the blind source separation problem, *In Proc. IEEE Int. Conf. on Acoustics, Speech, and Signal Processing,* pp. 2655–2658, Albuquerque, NM,, 1990

[15] E. Oja and J. Karhunen, Signal separation by nonlinear Hebbian learning, *In M. Palaniswami, Y. Attikiouzel, R. Marks, D. Fogel and T. Fukuda(Eds.), Computational Intelligence - A Dynamic System Perspective*, pp. 83–97, New York:IEEE Press, 1995.

[16] A. Hyvärinen and E. Oja, A fast fixed-point algorithm for independent component analysis, *Advances in Neural Information Processing Systems,* vol. 9, pp. 1483–1492, 1997.

[17] S. Miyamoto and M. Mukaidono, Fuzzy $c$-means as a regularization and maximum entropy approach, *Proc. 7th Int. Fuzzy Systems Association World Congress*, vol.2, pp. 86-92, 1997.

[18] K. Honda and H. Ichihashi, Linear fuzzy clustering techniques with missing values and their application to local principal component analysis, *IEEE Trans. on Fuzzy Systems*, vol.12, no.2, pp.183-193, 2004.

[19] F. Höppner, F. Klawonn, R. Kruse and T. Runkler, *Fuzzy Cluster Analysis*, Jhon Wiley & Sons, 1999.