Neural Comput & Applic (2007) 16: 167–172
DOI 10.1007/s00521-006-0058-6

ORIGINAL ARTICLE

Hongyi Peng · Siming Zhu

Handling of incomplete data sets using ICA and SOM in data mining

Received: 2 September 2005 / Accepted: 24 April 2006 / Published online: 30 May 2006

Abstract Based on independent component analysis (ICA) and self-organizing maps (SOM), this paper proposes an ISOM-DH model for the incomplete data's handling in data mining. Under these circumstances the data remain dependent and non-Gaussian, this model can make full use of the information of the given data to estimate the missing data and can visualize the handled high-dimensional data. Compared with mixture of principal component analyzers (MPCA), mean method and standard SOM-based fuzzy map model, ISOM-DH model can be applied to more cases, thus performing its superiority. Meanwhile, the correctness and reasonableness of ISOM-DH model is also validated by the experiment carried out in this paper.

Keywords Incomplete data · ICA (independent component analysis) · SOM (self-organizing maps) · Dependence · Non-Gaussian distribution

1 Introduction

Different from traditional statistical analysis, data mining is aimed at finding unsuspected relationships which are of value to the database owners or data miners. Due to the complex and huge volume of data, traditional methods cannot satisfy the need of data mining, so artificial intelligence techniques have been widely used in data mining. Among those artificial intelligence data mining methods, the self-organizing maps (SOM), proposed by Kohonen [1], have become one of the powerful techniques of data mining with cluster analysis by providing data visualization. It maps high-dimensional data onto low-dimensional pictures, usually in the form of two-dimensional map or one-dimensional map. Besides, it also allows the data miners to view the clusters. The SOM approach is superior to other cluster analysis methods in data mining in view of the power of data visualization for the data miners. And SOM's detail algorithm is shown in [2, 3].

The standard SOM approach, however, is not able to process incomplete data. On the other hand, it is a rare case where the data sets contain entries for all the variables and for each observation. Wang [2] provided an SOM-based fuzzy map model to handle incomplete data by assuming that the variables were mutually independent and then replaced all missing data with characteristic means. But the fact that dependence often exists among many variables makes it impractical for this model to be applied. Oba et al. [4] studied how to handle missing data with MPCA method and took into account the dependence of variables, but the premise was that every variable should submit to Gaussian distribution. Actually, many variables are non-Gaussian, so this method cannot be applied widely, either. In addition to the above, handling missing data was also addressed in [5–10]. Likewise, many different limitations were detected in all of them, so how to overcome these limitations is the main task of this work.

As a new statistic method, ICA develops in recent years. It aims to find a linear transformation for non-Gaussian data so as to make components to be independent as possible. ICA was proposed by Jutten and Herault [11] initially, and then a simplified ICA approach was provided by Singh and Rai [12]. In 2002, Kocsor and Csirik [13] studied Fast ICA and gave the code of Fast ICA algorithm. Later Theis [14] presented an ICA learning algorithm based on geometry method. The development of ICA makes it possible to make full use of the information of given data to estimate missing data.

This study is to add some important advantages to handling incomplete data sets. Considering the dependence between variables which do not submit to Gaussian distribution, the proposed ISOM-DH model in this paper is capable of obtaining visibility of high-dimensional data.

This paper is organized as following. Section 2 presents a brief introduction of ICA model and Fast ICA algorithm. Section 3 proposes a new model called as ISOM-DH to handle incomplete data, and the corresponding method called as ISOM-DH method. In

H. Peng (✉) · S. Zhu
Department of Applied Mathematics,
Sun Yat-sen University, Guangzhou 510275, China
E-mail: penghyi@yahoo.com.cn

Sect. 4, an experiment is carried out to demonstrate the proposed ISOM-DH method. Finally, the conclusion is provided in Sect. 5.

2 ICA model and Fast ICA algorithm

The general form of ICA model is as follows:

$$\mathbf{X} = \mathbf{AS} \tag{1}$$

where $\mathbf{X} = [X_1, X_2, ..., X_m]^{\mathrm{T}}$ and $\mathbf{S} = [S_1, S_2, ..., S_n]^{\mathrm{T}}$ are random vectors with zero mean and finite covariance, $\mathbf{S}$ has n statistically independent components, and $\mathbf{A}$ is a constant $m \times n$ matrix to be estimated.

In order to assure model (1) to be identifiable, the model should meet with the three requirements given below:

1. $\mathbf{S}$ is a vector with independent components, of which at most one is Gaussian.
2. $\mathbf{A}$ is a rectangular matrix with at most as many columns as rows, namely $1 \leq n \leq m$.
3. $\mathbf{A}$ is $m \times n$ full rank matrix.

The realization of Fast ICA algorithm is divided into three steps: (1) center the observations to make its mean zero, (2) whiten the observations and (3) extract the independent components. Steps (1) and (2) can be considered as pre-processing of observations, and they can simplify ICA algorithm.

In order to extract the independent components, the detail algorithm of step (3) is as follows:

1. Take a random initial vector $\mathbf{g}(0)$ of norm 1. Let $k = 1$.
2. Let $\mathbf{g}(k) = E\left\{\mathbf{X}\left(\mathbf{g}(k-1)^{\mathrm{T}}\mathbf{X}\right)^3\right\} - 3\mathbf{g}(k-1)$. Then the expectation can be estimated using a large sample of $\mathbf{X}$ vectors.
3. Divide $\mathbf{g}(k)$ by its norm.
4. If $\left|\mathbf{g}(k)^{\mathrm{T}}\mathbf{g}(k-1)\right|$ is not close enough to 1, let $k = k + 1$ and go back to step 2. Otherwise, output the vector $\mathbf{g}(k)$.

The final vector $\mathbf{g}(k)$ given by the algorithm is equal to one of the columns of the orthogonal mixing matrix $\mathbf{W} = \mathbf{A}^{-1}$. This means that $\mathbf{g}(k)$ separates one of non-Gaussian independent components.

To estimate n independent components, we run this algorithm n times. To ensure that we estimate each time a different independent component, we only need to add a simple orthogonal projection inside the loop. Note that the columns of the mixing matrix $\mathbf{W}$ are orthogonal because of the sphering. Thus, we can estimate the independent components one by one by projecting the current solution $\mathbf{g}(k)$ on the space orthogonal to the columns of the mixing matrix $\mathbf{W}$ previously found. Define the matrix $\bar{\mathbf{W}}$ as a matrix whose columns are the previously found columns of $\mathbf{W}$. Then add the projection operation at the beginning of step 3. Let $\mathbf{g}(k) = \mathbf{g}(k) - \bar{\mathbf{W}}\bar{\mathbf{W}}^{\mathrm{T}}\mathbf{g}(k)$. Divide $\mathbf{g}(k)$ by its norm.

Also the initial random vector should be projected this way before starting the iterations. To prevent the estimation errors in $\bar{\mathbf{W}}$ from deteriorating the estimate $\mathbf{g}(k)$, this projection step can be omitted after the first few iterations: once the solution $\mathbf{g}(k)$ has entered the basin of attraction of one of the fixed points, it will stay there and converge to that fixed point.

3 ISOM-DH model

ISOM-DH model is depicted in Fig. 1.

Firstly, the users need to confirm m indices, then sample appropriate complete data. Later, standardize the sampling data and choose n indices which are linearly independent. Refer to case 1 of Sect. 3.1 for the method of choosing indices. Assuming that n indices are mixed by n independent components, a data set of independent components can be got by ICA. By that the estimation of missing data can be carried out. Refer to Sect. 3.3 for the detailed process. Train the SOM and generate fuzzy map at last (cf. Sect. 3.4).

3.1 Choosing indices

For choosing indices, we discuss it in two cases:

Case 1:

Assume the rank of the covariance matrix of m indices is n ($n \leq m$), then n indices with linear independence can be found. The method for choosing the n indices is as follows:

Assume the size of sample is l, and the standardized indices are $X_1, X_2, ..., X_m$, where $X_i = (x_{i1}, x_{i2}, ..., x_{il})$ ($i = 1, ..., m$) (without causing any confusion, X_i can represent not only a random variable but also a row vector of samples). From the m indices, we can find n linearly independent indices, then let them be $X = (X_1, X_2, ..., X_n)^{\mathrm{T}}$ and let $C = XX^{\mathrm{T}}$. Usually the sample size l is much greater than m, so the storage space of the matrix C is much smaller than X. Suppose matrix $\bar{\mathbf{C}}$ consists of

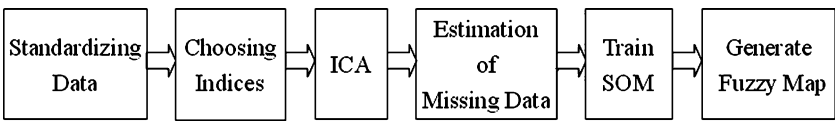

Fig. 1 ISOM-DH model

i_1th, i_2th, ..., i_nth rows and i_1th, i_2th, ..., i_nth columns from C, we notice that there exists corresponding relation between C and X, i.e., if the determinant of matrix $\bar{\mathbf{C}}$ is not equal to zero, then $X_{i_1}, X_{i_2}, \ldots, X_{i_n}$ is linearly independent. The detail algorithm of choosing indices is as follows:

1. Initialize $G = C$, $j = k$, $d[i] = 0$, $i = 1, ..., m$.
2. If det $(G) \geq \varepsilon_G$, then go to step 5, where det $(\cdot)$ represents the determinant of matrix, and ε_G is a predetermined small positive constant.
3. If det $(G) < \varepsilon_G$, then $d[j] = 1$, let G be the matrix which does not consist of the jth row and the jth column from G, and set $j = j - 1$.
4. If $j \geq 1$, then go to step 2, otherwise go to step 5.
5. Set $h = \sum_{i=1}^{m} d[i]$, if $d[j] = 0$, $j = 1, ..., m$, then X_j is the linearly independent row vector we need. And totally there are $(m - n)$ row vectors.

Case 2:

Let $\mathbf{X^{(c)}} = [X_1, X_2, \ldots, X_c]^{\mathrm{T}}$ be the missing data, $\mathbf{X^{(m-c)}} = [X_{c+1}, X_{c+2}, \ldots, X_m]^{\mathrm{T}}$ be the known data, $\mathbf{X} = \left(\mathbf{X^{(c)}}, \mathbf{X^{(m-c)}}\right)$, rank $(\mathbf{X}) = n$ and rank$\left(\mathbf{X^{(m-c)}}\right) = k$. Then k indices can be found from rank$\left(\mathbf{X^{(m-c)}}\right)$ according to the algorithm in case 1 and let them be $\mathbf{X^{(k)}}$. In addition, $(n-k)$ indices can be found from $\mathbf{X^{(c)}}$, let them be $\mathbf{X^{(n-k)}}$, and make Rank$\left(\mathbf{X^{(k)}}, \mathbf{X^{(n-k)}}\right) = n$. Let the correlation coefficient matrix of $\left(\mathbf{X^{(k)}}, \mathbf{X^{(c)}}\right)$ be $\mathbf{H}$, then the algorithm for finding the $\mathbf{X^{(n-k)}}$ is as follows:

1. Set $d[j] = 1$ $(j = 1, ..., k)$, $d[j] = 0$ $(j = k+1, ..., k+c)$ and tem $= k + 1$.
2. Set d[tem] $= 1$ and suppose matrix $\mathbf{M}$ consists of the 1st, ..., $(k+c)$th rows and the 1, ..., $(k+c)$ columns of $\mathbf{H}$. If det $(\mathbf{M}) < \varepsilon_G$, then d[tem] $= 0$, where det $(\cdot)$ represents the determinant of matrix, and ε_G is a predetermined small positive constant.
3. Set tem $=$ tem $+ 1$. If tem $> k + c$, then go to step 4, otherwise go to step 2.
4. If $d[j] = 1$, $(j = k+1, ..., k+c)$, then X_j is the row vector which we want. There are $(n - k)$ row vectors in sum and they compose the matrix $\mathbf{X^{(n-k)}}$.

3.2 The determination of linear correlation between indices

Let X_i, $(i = 1, ..., n)$, Y be the standardized data and suppose $X_1, X_2, ..., X_n$ are linearly independent. Let the correlation coefficient of Y and X_i be r_{Yi}. Let the correlation coefficient of X_i and X_j be r_{ij}. Assume $Y = k_1X_1 + k_2X_2 + \cdots + k_nX_n$.

For cov $(Y, X_i) = k_1$ cov $(X_1, X_i) + k_2$ cov $(X_2, X_i) + ...$ k_n cov (X_n, X_i), thus $r_{Yi} = k_1r_{1i} + k_2r_{2i} + \cdots + k_nr_{ni}$.

Their matrix form can be described as follows:

$$\mathbf{R}_{YX} = \mathbf{R}_{XX}\mathbf{K}$$

where $\mathbf{R}_{YX} = [\, r_{Y1}, ..., r_{Yn}\,]^{\mathrm{T}}$, $\mathbf{R}_{XX} = (r_{ij})_{n \times n}$, $\mathbf{K} = [k_1, ..., k_n\,]^{\mathrm{T}}$.

Since X_i, $i = 1, ..., n$ are linearly independent, we have that $\mathbf{R}_{XX}$ is inverse. Thus

$$\mathbf{K} = \mathbf{R}_{XX}^{-1}\mathbf{R}_{YX}$$

Therefore, k_i, $i = 1, ..., n$ can be determined.

3.3 The estimation of missing data

From the initial data set, we can sample appropriate complete data. Then a data set of independent components can be obtained by ICA. The probability density function (pdf) of every independent component can be estimated by support vector machine (SVM). Refer to [15] for the detail process of estimation of probability density function.

Assume the density function $p(S_i)$, $(i = 1, ..., n)$ of every independent component has been estimated. Thus the joint pdf of independent component vector is $p(\mathbf{S}) = \prod_{i=1}^{n} p(S_i)$, where $\mathbf{S} = [S_1, S_2, ..., S_n]^{\mathrm{T}}$.

Let $\mathbf{X}^*$ be the original data, $\mathbf{X}$ be the standardized data, and suppose rank $(\mathbf{X}) = n$. Assume $\mathbf{X}^{(c)} = [X_1, X_2, ..., X_c\,]^{\mathrm{T}}$ and $\mathbf{X}^{(m-c)} = [X_{c+1}, X_{c+2}, ..., X_m]^{\mathrm{T}}$ are the missing part and observed part in $\mathbf{X}$ respectively. To explain how to estimate the missing part $\mathbf{X}^{(c)}$, the following different cases are analyzed and discussed.

Case 1: $m = n$.

We have $\mathbf{X} = \tilde{\mathbf{B}}\tilde{\mathbf{X}} = \tilde{\mathbf{B}}\mathbf{W}\mathbf{S} = \bar{\mathbf{B}}\mathbf{S}$, where $\tilde{\mathbf{X}}$ denotes the whiten data vector, $\mathbf{S} = [S_1, S_2, ..., S_n]^T$ denotes the independent vector, $\bar{\mathbf{B}} = \left(\bar{b}_{ij}\right)_{n \times n}$, both $\bar{\mathbf{B}}$ and $\mathbf{W}$ are $n \times n$ matrixes. According to Sect. 3.1, c independent components which can be found are linearly independent with the observed part $\mathbf{X}^{(m-c)}$, and let them be S_i, $i = 1, ..., c$. According to the property of pdf, we have:

$$p(S_1, \ldots, S_c, X_{c+1}, \ldots, X_n) = \mathbf{D}p(S_1, \ldots, S_n)$$

where $\mathbf{D}$ is absolute value of Jacobian matrix determinant. Thus we have:

$$p(S_i|X_{c+1}, \ldots, X_n) = \frac{\int \cdots \int p(S_1, \ldots, S_n)\mathrm{d}S_1 \cdots \mathrm{d}S_{i-1}\mathrm{d}S_{i+1} \cdots \mathrm{d}S_c}{\int \cdots \int p(S_1, \ldots, S_n)\mathrm{d}S_1 \cdots \mathrm{d}S_c}$$

where $1 \leq i \leq c$. Thus we have the S_i's estimation :

$$\hat{S}_i = \int S_i p(S_i|X_{c+1}, \ldots, X_m)\mathrm{d}S_i, \quad 1 \leqslant i \leqslant c$$

According to Sect. 3.2, $\hat{S}_k (c < k \leq n)$ can be linearly expressed by $\hat{S}_1, \ldots, \hat{S}_c, X_{c+1}, \ldots, X_n$. Thus we have standardized missing part's estimation:

$$\hat{X}_i = \bar{b}_{i,1}\hat{S}_1 + \bar{b}_{i,2}\hat{S}_2 + \cdots + \bar{b}_{i,n}\hat{S}_n, \quad i = 1, \ldots, c$$

Combined with expectation and variance of the original missing part, the original missing part can also be estimated.

Case 2: $1 \leq n < m$.

Let rank $(\mathbf{X}^{(m-c)}) = k$, $1 \leq k \leq \min(m-c, n)$ and suppose $\Re(\mathbf{X})$ denotes the linear space of the row vectors of $\mathbf{X}$.

1. If $k = n$, then $\Re(\mathbf{X}^{(c)}) \subseteq \Re(\mathbf{X}^{(m-c)})$. According to Sect. 3.2, X_i $(i = 1, ..., c)$ can be linearly expressed by X_i, $i = c+1, ..., m$. Thus the original missing data can be estimated with the expectations and variances of the original missing data.
2. If $k < n$, then there are k linear independent row vectors in $\mathbf{X}^{(m-c)}$. According to Sect. 3.1, we can find the k linear independent row vectors and let them be $\mathbf{X}^{(k)} = [X_{c+1}, ..., X_{c+k}]^{\mathrm{T}}$. Because $\Re(\mathbf{X}) = n$, we can find $(n-k)$ linear independent row vectors, and let them be $\mathbf{X}^{(n-k)} = [X_1, ..., X_{n-k}]^{\mathrm{T}}$. Let $\mathbf{X}^{(n)} = [\mathbf{X}^{(k)}\mathbf{X}^{(n-k)}]$, then $\Re(\mathbf{X}^{(n)}) = \Re(\mathbf{X}) = n$. Likewise, the estimation of missing part $\mathbf{X}^{(n-k)}$ is similar to case 1. Let $\mathbf{X}^{(c-k)} = [X_{n-k+1}, ..., X_c]^{\mathrm{T}}$, then $\Re(\mathbf{X}^{(c-k)}) \subseteq \Re(\mathbf{X}^{(n-k)})$. According to Sect. 3.2, we can estimate the missing part $\mathbf{X}^{(c-k)}$ too. Finally, combined with their means and variances, all original missing parts can be estimated.

3.4 Generate the fuzzy map

After the estimation of missing data, we can obtain a fuzzy data set. Based on the fuzzy data set, we can get a corresponding fuzzy data set of independent components. Assume we have trained the SOM, then we can generate the fuzzy map, and the detail process is as follows:

Step 1 Create a two-dimensional coordinate. The horizontal axis **J** has N bins arranged in the order of the N output nodes of the SOM. The vertical axis **A** represents the strength of the generated clusters by the trained SOM.
Step 2 For the independent components $\mathbf{S}^{(i)}$ of each observation $\mathbf{X}^{(i)}$ $(i = 1, ..., l)$, perform the following substeps.
Step 2.1 Present $\mathbf{S}^{(i)}$ to the trained SOM. It is then organized by the trained SOM to the output node j $(j = 1, ..., N)$.
Step 2.2 Locate bin j on the **J** axis. On the **A** direction of bin j, add a gray bar (with height = 1) for the non-fuzzy observation, and plus a black bar (with height = 1) on the top for fuzzy observation.

4 Experimental results

In this section, we use a real-world example to demonstrate the method described. The data sets of main economic indicators for major retail enterprises from Beijing Statistic Annual (2002, 2003) are used to test the model proposed in this paper. We sample 290 preliminary observations with six economic indices, i.e., X1, X2, X3, X4, X5 and X6 represent sales revenue, total profits, sales value per capita, pre-tax per capita, ratio of pre-tax to sales revenue and turnover rate of stock (%), respectively. If the sample data size is less than 2,000, we use Shapiro–Wilk W method to test if every variable is Gaussian or not, and the results show that the six variables are non-Gaussian distribution. Then we apply ICA to the randomly selected 230 complete observations. Then we, respectively, sample 60, 70, 80 and 100 observations from the 290 complete observations at random. Let X2 and X4 be null. Thus, we get an incomplete data set which has 290 samples. Then the missing data and independent components of every observation can be estimated according to Sects. 3.1, 3.2 and 3.3. Then we use ISOM-DH for the independent components of the incomplete data sets. The equipment is a PC with Windows XP operating system and SAS9.0 Software. At last, we use means method, MPCA method, ISOM-DH method, respectively, to estimate missing data, and get the corresponding residuals of X2 and X4. The case of sampling 100 samples is displayed in this paper.

The scatter plots of residuals of X2 and X4 are displayed in Figs. 2 and 3, respectively. Number is the order of incomplete observations; XPR2, PCASR2 and ICASR2 are the residuals of X2 by mean, MPCA and ISOM-DH methods respectively; XPR4, PCASR4 and ICASR4 are the residuals of X4 by mean, MPCA and ISOM-DH methods, respectively. The residual square means of three methods are stated in Table 1. From Figs. 2, 3 and Table 1, we know that the precision of estimating missing data by ISOM-DH method is clearly prior to that by mean and MPCA methods.

The SOM topology is set to 100 output nodes, 100 nodes for the initial neighborhood, the initial learning rate of 0.85; 1 node for the final neighborhood, the final learning rate of 0.0001; and 2,320 learning iterations.

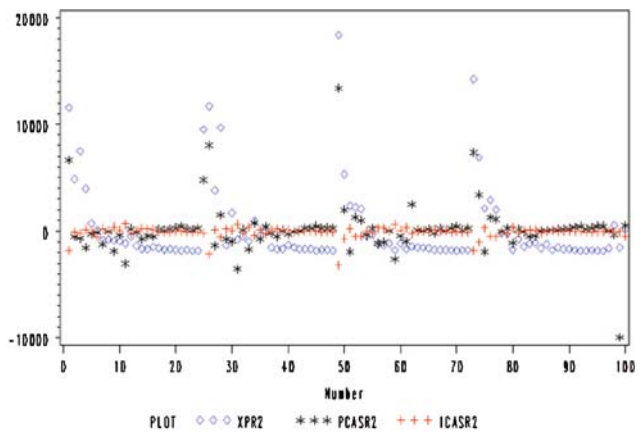

Fig. 2 The residuals of X2

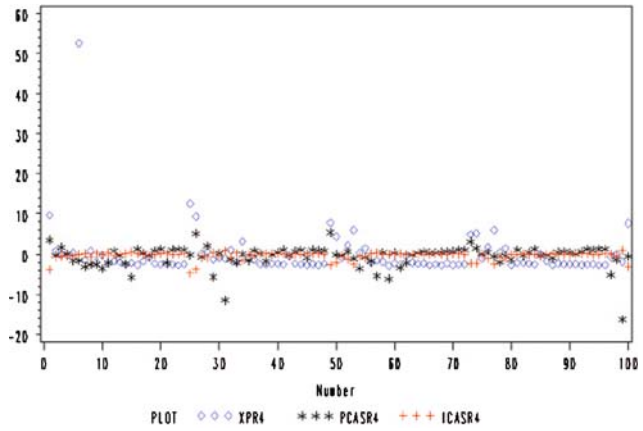

Fig. 3 The residuals of X4

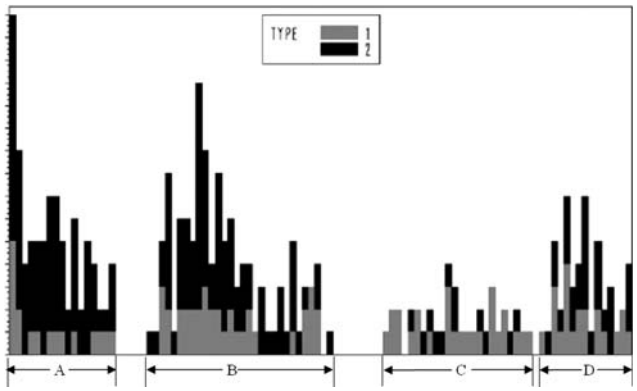

Fig. 4 The map results of ISOM-DH

Table 1 Comparison of the results sampling 100 observations

Indicator	Method	Mean of residual square
X2	Means	13,789,596.57
	MPCA	5,536,853.42
	ISOM-DH	9,440.15
X4	Means	37.3312510
	MPCA	7.7899376
	ISOM-DH	0.0016714

Figure 4 shows the trial result of ISOM-DH map. In Fig. 4, type = 1 represents the data records without missing data, and type = 2 represents the data records with missing data. Four clusters (marked A, B, C and D, respectively in Fig. 4) are identified on the ISOM-DH map. Among the four clusters, A, B and D are relatively clear, while C is fuzzy. Cluster D is contributed mainly by the incomplete observations.

The original multivariate data corresponding to the clusters are extracted. The characteristics of every cluster are depicted in Table 2. The characteristics of data set are depicted in Table 3. Compared with indicator means of data set in Table 3, the corresponding means of indicators for cluster A in Table 2 are much less. Similar to the above, the mean of indicator X6 for cluster B in Table 2 is greater than that in Table 3, the means of indicators X1, X2, X3, X4 for cluster C in Table 2 are greater than that in Table 3, and the mean of indicator X1 for cluster D in Table 2 is much greater than that in Table 3. By that we can easily and clearly understand each cluster in Fig. 4.

5 Conclusion

In data mining, SOM is considered a useful technique for it allows the data miners to view high-dimensional data in the data mining field. However, the standard SOM approach is not able to process incomplete data, and it is also a rare case where the data sets contain entries for all of the variables for each observation. SOM-based fuzzy map model for data mining was also proposed to deal with incomplete data [2], but it cannot apply to cases where there is dependence between data. Here we have presented the ISOM-DH model for the handling of incomplete data sets and it has some advantages over the previous methods. First, ISOM-DH has better precision in estimating missing data than mean and MPCA methods. Second, since ISOM-DH model takes into account the dependence of data, it can be applied to more cases than SOM-based fuzzy map model. ISOM-DH model has one key component: handing missing data by ICA. The results in the experiment have demonstrated that SOM can be more useful in data mining with incomplete data if the proposed ISOM-DH is applied.

Data mining is a growing field in the information era, and the original data sets for data mining are usually incomplete. The ISOM-DH model provides an effective data mining technology, so that the data miners can make full use of the information of observations with missing data to verify and strengthen the data mining results derived from limited observations with only complete data.

Table 2 Characterizations of every cluster

Cluster	Indicator					
	X1	X2	X3	X4	X5	X6
A						
Mean	8,860.7865	286.4719	15.2079	0.4820	3.2416	1,081.3539
SD	6,341.0221	244.6665	5.8136	0.2691	1.1432	1,755.1082
B						
Mean	14,330.2957	1,303.800	34.8217	3.1626	10.3678	4,692.8774
SD	17,131.7244	1,840.9455	29.0415	3.9818	15.7727	7,583.5040
C						
Mean	93,725.8158	8,470.1316	84.6763	8.5289	9.5789	1,385.7342
SD	27,765.1374	2,859.6255	36.3026	5.5195	5.9264	693.6077
D						
Mean	43,067.4792	1,272.2708	46.9875	1.2750	2.8250	1,581.8625
SD	16,232.1752	964.5904	20.7722	0.9562	1.5098	1,209.3315

Table 3 Characterizations of data set

Indicator	X1	X2	X3	X4	X5	X6
Mean	27,811.7724	1,925.4034	37.3486	2.7307	6.8289	2636.2369
SD	32,594.6706	3,348.7017	39.7925	5.4606	10.7705	17,440.8399

References

1. Kohonen T (1982) Self-organized formation of topologically correct feature maps. Biol Cybern 43:59–69
2. Wang S (2003) Application of self-organising maps for data mining with incomplete data sets. Neural Comput Appl 12:42–48
3. Chang P-C, Lai C-Y (2005) A hybrid system combining self-organizing maps with case-based reasoning in wholesaler's new-release book for forecasting. Expert Syst Appl 29:183–192
4. Oba S et al (2002) Missing value estimation using mixture of PCAs. LNCS 2415, pp 492–497
5. Ad Feelders (1999) Handling missing data in trees-surrogate splits or statistical imputation. LNAI 1704, pp 329–334
6. Grzymala-Busse JW (2004) Rough set approach to incomplete data. LNAI 3070, pp 50–55
7. Gerardo BD et al (2004) The association rule algorithm with missing data in data mining. LNCS3043, pp 97–105
8. Li D et al (2004) Towards missing data imputation—a study of fuzzy K-means clustering method. LNAI 3066, pp 573–579
9. Zs. J. Viharos et al (2002) Training and application of artificial neural networks with incomplete data. LNAI 2358, pp 649–659
10. Latkowski R (2002) Incomplete data decomposition for classification. LNAI 2475, pp 413–420
11. Jutten C, Herault J (1998) Independent component analysis versus PCA. In: Proceeding of European signal processing conference, 287–314
12. Singh Y, Rai CS (2003) A simplified approach to independent component analysis. Neural Comput Appl 12:173–177
13. Kocsor A, Csirik J (2001) Fast independent component analysis in kernel feature spaces. LNCS 2234, pp 271–281
14. Theis FJ et al (2002) Overcomplete ICA with a geometric algorithm. LNCS 2415, pp 1049–1054
15. Vapnik V (2004) Statistical learning theory. Publishing House of Electronics Industry, Beijing