Classification Mapping and Its Application on Chemical Systems

Hang Su,*,† Zhi-Hong Che,‡ Jian-Min Wu,† and Rong Li†

Department of New Material Research, Central Iron and Steel Research Institute, Xueyan South Road, 100081 Beijing, China, and Fourth Department of Second Academy, Aerospace Industrial Corporation, 100854 Beijing, China

Received January 4, 1999

In this paper, the conception and the construction of classification maps (CMs) are discussed. Methods called Fp mapping and Sp mapping are proposed based on discriminant analysis (DA) and projection pursuit (PP). By these methods, an experimental data set with classification information can be mapped from multidimensional data space onto a series two-dimensional projection keeping the original classification structure in the process. These projections show intuitively the semiquantitative relationships between the study object and its controlling factors that we cannot observe directly from multivariate data space. Based on existing experimental data, these methods are useful in giving valuable suggestions on further experimental works. As an example, they were applied to the study of a kind of propellant material and a kind of welding material and were proven effective in the optimization of the material's performance.

1. INTRODUCTION

In the past 10 years, pattern recognition (PR) methods were applied to solve some experimental and theoretical problems of chemical systems and show their powerful function in these fields.[1−4,7,11] Some of these works touched upon the technique of projection pursuit (PP) which was developed from multianalysis. It is applied to project a multidimensional data space onto two-dimensional planes and keep some characteristics constantly in that process. The key of PP is to define a projection index and then find a suitable algorithm to accomplish the projection.[5]

Suppose there is an N-dimensional data set that classified several classes by a certain characteristic. We may get a group of projection maps by mapping these data points from the N-dimensional space onto two-dimensional planes and expecting some of them to show us some further information (e.g. the classification structure) of the given data set. Theoretically speaking, any space not lower than three-dimension may give two-dimensional projections infinitely, but only a few of them can show regularities that we expected. We call projections with the following qualities as classification maps (CMs): (1) The classification structure of the original data space can be revealed on these projections. That makes data points of different classes can be shown separately on these projections. (2) There exists a one-to-one relationship between the original space and its projections. That means a point in the original data space should be able to be calculated uniquely from its projected two-dimensional points by an inverse mapping process. The mapping from which the CMs are constructed is named as classification mapping.

The classification mapping gives qualitative or semiquantitative regularities of the study object. It is suitable to systems that have some original study results, but for the following reasons there are some difficulties on further researching. (1) The system has too many controlling factors and is too complex to analysis. (2) There are inexplainable mechanisms in studies. (3) The system has strong nolinearity and is difficult to be modeling strictly.

In this paper, we will first discuss some groundwork problems of CMs, then put forward the theoretical discussion of Fp mapping and Sp mapping, and next bring out the general procedure of CMs construction from a given data set. Examples in application are given at the end of this paper.

2. THEORETICAL GROUNDWORKS

Problem. Suppose the characteristic $\boldsymbol{E}$ of a chemical system **P** are interrelated with N features $x_1, x_2, ..., x_N$. The n given samples $P_i = (x_{i1}, x_{i2}, ..., x_{iN})$ $(i = 1, 2, ..., n)$ are divided into two classes Ω_+ and Ω_- by the measurement of $\boldsymbol{E}$. For example

$$P_i \in \Omega_+ \quad \text{if } \boldsymbol{E} > E_0$$

$$P_i \in \Omega_- \quad \text{if } \boldsymbol{E} \le E_0$$

Here E_0 is a preset threshold value. The problem is how can we construct CMs for such a sample set?

2.1. About the Principal Component Analysis. The principal component analysis (PCA)[6] was used to construct CMs by some author.[2,7] Being a linear mapping, PCA gives a brief functional relationship between point $(x_1, x_2, ..., x_N)$ on the original data space and its imaged point $(y_1, y_2, ..., y_N)$ on the projected maps.

$$\mathbf{x} = \mathbf{U}^{-1}\mathbf{y}$$

* E-mail: hhome@ihw.com.cn.
† Central Iron and Steel Research Institute.
‡ Aerospace Industrial Corporation.

10.1021/ci990302o CCC: $18.00
Published on Web 06/03/1999

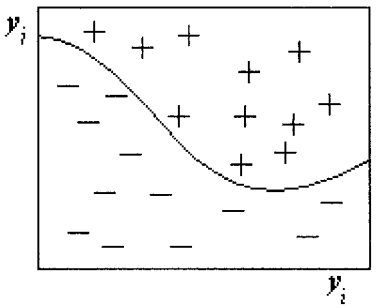

Figure 1. An idealized projection map constructed from PCs y_i and y_j. Symbols "+" and "−" denote the classified samples.

Here $\mathbf{y} = (y_1, y_2, ..., y_N)^T$,$\mathbf{x} = (x_1, x_2, ..., x_N)^T$, and **U** is the eigenvector matrix of the covariance matrix **S** (**S** = $\mathbf{X^TX}$, where **X** is the given data matrix). [In a general way, only 2−3 PCs are taken to construct CMs. Therefore, additional information is needed frequently to accomplish an inverse mapping process. For example, if we design a point on a two-dimensional map constructed of y_i and y_j, then the coordinate of other y_k ($k = 1, 2, ..., N, k \neq i, j$) are often set as the center of the optimal class

$$y_k = \sum_{i=1}^{n} y_{ik} | P_i \in \Omega_+$$

It is necessary to get back from a projected point to its original point in space uniquely.]

The typical process of its construction can be summarized as follows.[7]

Step 1. Classify the given *n* samples by the measurement ***E***.

Step 2. Calculate the principal component vectors of the given data matrix **X**.

$$\mathbf{Y} = \mathbf{U}\,\mathbf{X}$$

Here columns **Y** are the principal component vectors of matrix **X**.

Step 3. Mark the *n* samples onto the projection map with axes y_i and y_j. If the two principal components (PCs) y_i and y_j are selected properly, we may get a projection map like Figure 1.

If a projection map given by PCA shows regularities on data distribution, it should be useful in many applications. Reference 7 shows an application of PCA and its inverse mapping on the material design for high-T_c superconductors.

In many cases, however, PCA mapping cannot give practicable CMs although the samples are classable in space. The reasons are as follows.

a. PCA method does not relate to the classificatory information of a sample set in principle. Theoretically speaking, the first and the second PCs covered the most information of the given data space, but that may not be the criterion to construct CMs (see Figure 2). In fact, some practicable projections were constructed from other less important PCs in ref 11 and our applications.

b. Being a linear mapping, PCA method cannot separate sample sets with "included type" distribution (see Figure 3).

c. In the third step above, the "proper" projection map is given by observation factitiously. Considering that PCA method gives *N* principal components, which may construct

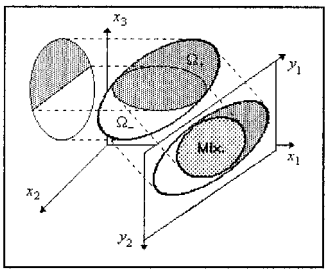

Figure 2. PCA method does not relate to the classificatory information in principle. Notice that the direction of PCs y_1 and y_2 are not the same as that of the classification (x_3).

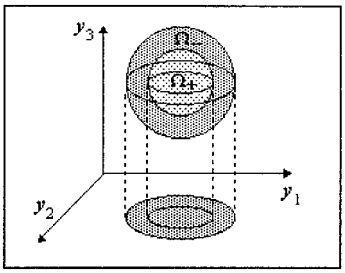

Figure 3. Linear mapping like PCA cannot separate a sample set with "included type" distribution.

$N(N-1)/2$ projections by twos, it seems rather indefinite to ascertain one or more available maps from those projections only by the unaided eye.

In short, the PCA method is not the seemly way for CMs construction. Proper methods can be found by taking into account the discriminant analysis.

2.2. About the Discriminant Analysis (DA). Discriminant analysis is a subfield of pattern recognition. Here we give only a brief introduction on some related methods. References 8−10 have the detail.

2.2.1. The Linear Discriminant Function (LDF). The general form of LDF is

$$F = \sum_{j=1}^{N} a_j x_j + a_0 \tag{1}$$

It makes the following discriminant have a minimized error on discrimination to the given sample set $P_i = (x_{i1}, x_{i2}, ..., x_{iN})$ ($i = 1, 2, ..., n$)

$$\begin{cases} P_i \in \Omega_+ & \text{when } F(P_i) < 0 \\ P_i \in \Omega_- & \text{when } F(P_i) \geq 0 \end{cases}$$

The classification error of a discriminant function is often estimated approximately by an error-counting procedure.[10] The error expression is

$$\text{Ed\%} = \frac{\text{the number of miss classified samples}}{\text{the number of total samples}} \times 100\% \tag{2}$$

where the numerator means the number of samples that has been misclassified by the discriminant function. A typical algorithm that gives an optimal LDF statistically is the Fisher method.[8]

$$\begin{cases} \mathbf{A} = (\mathbf{S}_+ + \mathbf{S}_-)^{-1}(\bar{\mathbf{X}}_+ - \bar{\mathbf{X}}_-) \\ a_0 = -\dfrac{1}{n}(n_+ \mathbf{A}^T \bar{\mathbf{X}}_+ + n_- \mathbf{A}^T \bar{\mathbf{X}}_-) \end{cases} \tag{3}$$

Here $\mathbf{A} = (a_1, a_2, ..., a_N)$, $\mathbf{S}_+$, $\mathbf{S}_-$ is the covariance matrix of Ω_+ and Ω_-, respectively, and $\mathbf{S} = \mathbf{X}^T\mathbf{X}$, $\bar{\mathbf{X}}_+$,$\bar{\mathbf{X}}_-$ is the mean value of Ω_+ and Ω_-, respectively. Here

$$\bar{\mathbf{X}} = \frac{1}{n}\sum_{i=1}^{n} P_i$$

n_+, n_- is the number of Ω_+ and Ω_-, respectively.

2.2.2. The Spheroidic Discriminant Function (SDF). The general form of SDF is

$$F = \sum_{j=1}^{N}(x_j - x_j^0)^2 - R^2 \tag{4}$$

And to any sample $P_i = (x_{i1}, x_{i2}, ..., x_{iN})$ $(i = 1, 2, ..., n)$, there are

$$\begin{cases} P_i \in \Omega_+ & \text{when } F(P_i) < 0 \\ P_i \in \Omega_- & \text{when } F(P_i) \geq 0 \end{cases}$$

A typical algorithm to give SDF is the Hamming method.[9] In this paper, it is extended to fit data set with spheroid like classification structures. Suppose we concern with samples belong to Ω_+.

Step 1. Process the following whitening transformation[10]

$$\mathbf{X}^{(\text{new})} = \Lambda_+^{-1/2}\mathbf{U}_+\mathbf{X}^{(\text{old})}$$

The aim of such a transformation is to glomerate the data of Ω_+ to a normal spherical distribution. Here Λ is the diagonal matrix of eigenvalues of the covariance matrix $\mathbf{S}$ $(\mathbf{S} = \mathbf{X}^T\mathbf{X})$

$$\Lambda = \begin{bmatrix} \lambda_1 & & & 0 \\ & \lambda_2 & & \\ & & \vdots & \\ 0 & & & \lambda_N \end{bmatrix}$$

The Λ_+ is given by solving the following eigenvalue problem:

$$\mathbf{S}_+\mathbf{U}_+ = \Lambda_+\mathbf{U}_+$$

$\mathbf{S}_+$ is the covariance matrix of Ω_+, and $\mathbf{U}_+$ is the eigenvectors matrix of $\mathbf{S}_+$.

Step 2. Calculate the Hamming center

$$C^0 = (x_1^0, x_2^0, ..., x_N^0)$$

and

$$x_j^0 = \frac{1}{n_+}\sum_{i=1}^{n_+} x_{ij}|_{P_i \in \Omega_+}$$

Step 3. Calculate the distance from each samples to the Hamming center

$$d(P_i, C^0) = [\sum_{j=1}^{N}(x_{ij} - x_j^0)^2]^{1/2}$$

Step 4. Sort ascending $d(P_i, C^0)$ and find a threshold value of R that makes the following discriminant have a minimum Ed% (see formula 2) for all given samples.

$$\begin{cases} P_i \in \Omega_+ & \text{when } d(P_i, C^0) < R \\ P_i \in \Omega_- & \text{when } d(P_i, C^0) \geq R \end{cases}$$

Steps 2−4 show the process of seeking an SDF. It is based on an improved Hamming method. Genetic algorithms (GAs) can be applied to search for an SDF also. Here gives a simplified definition.

The $N + 1$ parameters, including the N coordinate parameters of the center point $x_1^0, x_2^0, ..., x_N^0$ and the radius R, are considered as the encoding object in genetic algorithms. The ranges of these parameters are

$$x_j^0 \in [x_j^a, x_j^b] \quad (j = 1, 2, ..., N)$$

$$R \in [0, R_{\max}]$$

$R_{\max}$, x_j^a, and x_j^b should be defined according to circumstances, while in the following example (see section 4.2.2) they are defined as

$$R_{\max} = [\sum_{j}^{N}(x_j^b - x_j^a)^2]^{1/2}$$

$$x_j^a = x_j(\min) - [x_j(\max) - x_j(\min)]/2$$

$$x_j^b = x_j(\max) + [x_j(\max) - x_j(\min)]/2$$

$x_j(\max)$ and $x_j(\min)$ are the maximum and minimum value of x_j in the given data set. The fitness function can be defined as

$$\text{Fitness} = -\text{Ed}\%$$

For the principles of GAs please see related refs 13 and 14.

2.3. The Classification Mapping. As we know that, the monotone function and the peaked function are the two basic function types in modality. The topological structures of classified sample sets can be abstracted into two basic types correspondingly: the partitioned type and the included type (see Figure 4). A partitioned type sample set has an opened interface between the two classes, while an included type sample set has an enclosed interface between the two classes. Other classification structures that are more complicated can also be related with the two basic function types (Figure 5 shows some examples). The Fp mapping and the Sp mapping are proposed for the two basic classification structures respectively.

2.3.1. The Fp Mapping. The Fp mapping, which is suitable for data sets with partitioned type classification structures, is based on the LDF.

(1) Definition and Characteristics of Fp Mapping. Definition. Known that the N-dimensional data space $\mathbf{R}^N$ is constructed by N features $x_1, x_2, ..., x_N$, and the LDF of sample set with two classes Ω_+, Ω_- is known as

$$F = \sum_{i=1}^{N} a_i x_i + a_0 \tag{1}$$

Construct an orthogonal multinomial for each x_i

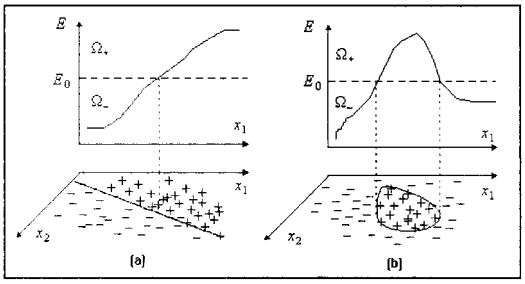

Figure 4. The two basic function types and its corresponding topological structures of classification: (a) monotone function and the corresponding partitioned type classification structure and (b) peaked function and the corresponding included type classification structure.

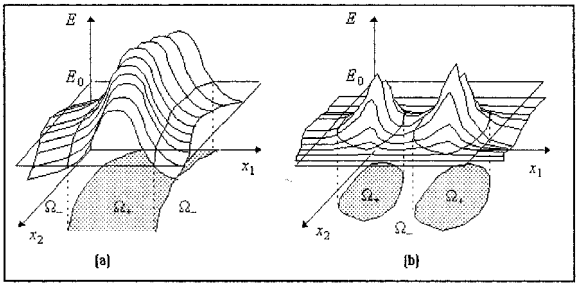

Figure 5. Complicated classification structures can be also attributed to the two basic function types: (a) sandwich type classification can be formed by a peaked function (x_1) and a monotone function (x_2) and (b) multioptimized-zone type classification can be formed by multipeaked functions.

$$y_i = \frac{1}{\phi_i} \sum_{\substack{j=1 \\ j \neq i}}^{N} a_j x_j \tag{5}$$

where ϕ_i is a normalization factor

$$\phi_i = [\sum_{\substack{j=1 \\ j \neq i}}^{N} a_j^2]^{1/2} \quad (i = 1, 2, ..., N) \tag{6}$$

Denote the plane constructed by x_i, y_i as $\boldsymbol{P}x_i,y_i$. Then the mapping from $\mathbf{R}^N$ to $\boldsymbol{P}x_i,y_i$ is called as a Fp mapping that formed by x_i. It is denoted as

$$\text{Fp: } \mathbf{R}^N \rightarrow \boldsymbol{P}x_i,y_i \quad (i = 1, 2, ..., N) \tag{7}$$

It has the following characteristics:

Characteristic 1. The hyperplane $F = 0$ in $\mathbf{R}^N$ is projected to a line $f_i = 0$ on plane $\boldsymbol{P}x_i,y_i$

$$f_i = \phi_i y_i + a_i x_i + a_0 \quad (i = 1, 2, ..., N) \tag{8}$$

Characteristic 2. Point (x_i^*, y_i^*) on plane $\boldsymbol{P}x_i,y_i$ correspond with point $(x_1, x_2, ..., x_N)$ in space $\mathbf{R}^N$ one-to-one, and the relationship between them is

$$x_k = \begin{cases} \left(\dfrac{\phi_i y_i^* - \sum_{\substack{j=1 \\ j \neq i}}^{N} a_j x_j^0}{\phi_i^2} \right) a_k + x_k^0 & k \neq i \\ x_i^* & k = i \end{cases} \quad (k = 1, 2, ..., N) \tag{9}$$

Here $(x_1^0, x_2^0, ..., x_N^0)$ is a reference point.

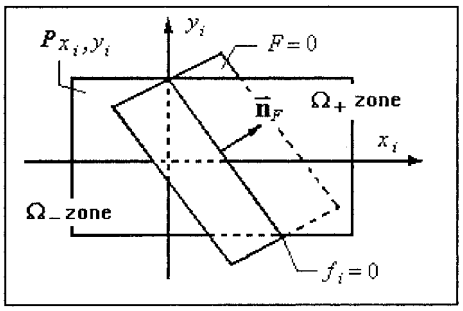

Figure 6. The geometrical meaning of Fp mapping.

(2) Related Derivation and Discussion. Formula 8 is but the reform of formulas 1 and 5. Nevertheless, it indicates that the projection map $\boldsymbol{P}x_i,y_i$ shows directly the classification structure of the N-dimensional data space as that defined by the LDF. In other words, if two imaged data points on plane $\boldsymbol{P}x_i,y_i$ are separated by line $f_i = 0$, then their inverse-image point in space $\mathbf{R}^N$ must be also separated by hyperplane $F = 0$. The PCA mapping has no such characteristic.

Characteristic 2 gives the inverse mapping from $\boldsymbol{P}x_i,y_i$ to $\mathbf{R}^N$. The derivation is given below. On the plane that is determined by axis x_i and reference point $(x_1^0, x_2^0, ..., x_N^0)$, the parametric function of axis y_i is

$$\begin{cases} \dfrac{x_j - x_j^0}{a_j} = t \quad (j = 1, 2, ..., N, j \neq i) \\ x_i = 0 \end{cases}$$

which gives

$$x_j = a_j t + x_j^0 \quad (j \neq i) \tag{10}$$

Substitute (10) for (5), we then have

$$y_i = \frac{1}{\phi_i} \sum_{\substack{j=1 \\ j \neq i}}^{N} (a_j^2 t + a_j x_j^0) = \frac{\phi_i^2 t + \sum_{\substack{j=1 \\ j \neq i}}^{N} a_j x_j^0}{\phi_i}$$

Therefore,

$$t = \frac{\phi_i y_i - \sum_{\substack{j=1 \\ j \neq i}}^{N} a_j x_j^0}{\phi_i^2} \tag{11}$$

when (x_i^*, y_i^*) is given, substitute t for (10), and the corresponded inverse-image expression (9) is acquired.

In a general way, the initial value of the reference point $(x_1^0, x_2^0, ..., x_N^0)$ is suggested to locate at the center of the optimal class

$$x_j^0 = \frac{1}{n_+} \sum_{i=1}^{n_+} x_{ij}|_{P_i \in \Omega_+} \quad (k = 1, 2, ..., N) \tag{12}$$

While a point (x_i^*, y_i^*) is designed on Fp maps and the inverse-image point $(x_1, x_2, ..., x_N)$ is calculated by formula

9, the reference point can be moved to $(x_1, x_2, ..., x_N)$ for the next design. This process can be accomplished easily by a computer program (vide infra).

(3) The Geometrical Meaning of Fp Mapping. The N-dimensional space $\boldsymbol{R}^N$ is divided into two region by a hyperplane $F = 0$. $\vec{\mathbf{n}}_F$ is the normal vector of the hyperplane. The $\boldsymbol{P}x_i,y_i$ paralleled with $\vec{\mathbf{n}}_F$, it is an orthogonal section of the hyperplane $F = 0$. $f_i = 0$ is the line of intersection of planes $\boldsymbol{P}x_i,y_i$ and $F = 0$ (see Figure 6). The N orthogonal projections $\boldsymbol{P}x_i,y_i$ contain integrally information of the given data space.

2.3.2. The Sp Mapping. The Sp mapping, which is suitable for data sets with included type classification structure, is based on the SDF. Being nonlinear mapping, there are some complexities on derivation.

(1) Definition and Characteristics of Sp Mapping. Definition. It is known that the N-dimensional data space $\mathbf{R}^N$ is constructed by N features $x_1, x_2, ..., x_N$, and the SDF of sample set with two classes Ω_+, Ω_- is known as

$$F = \sum_{j=1}^{N}(x_j - x_j^0)^2 - R^2 \tag{4}$$

$C^0 = (x_1^0, x_2^0, ..., x_N^0)$ is the center of the sphere $F = 0$. Suppose there is a known *reference point* $C' = (x_1', x_2', ..., x_N')$ and its symmetrical point about C^0: $C'' = (2x_1^0 - x_1', 2x_2^0 - x_2', ..., 2x_N^0 - x_N')$ $(C^0 \neq C' \neq C'')$. Construct a nonlinear item r_i for each x_i

$$r_i = \begin{cases} [\sum_{\substack{j=1 \\ j\neq i}}^{N}(x_j - x_j^0)^2]^{1/2} + r_i^0 & \text{when } r_i' \leq r_i'' \\ -[\sum_{\substack{j=1 \\ j\neq i}}^{N}(x_j - x_j^0)^2]^{1/2} + r_i^0 & \text{when } r_i' > r_i'' \end{cases} \tag{13}$$

where the r_i^0 is a *shift factor* and r_i' and r_i'' are *reference factors*

$$r_i^0 = [\sum_{\substack{j=1 \\ j\neq i}}^{N}(x_j^0)^2]^{1/2} \quad r_i' = [\sum_{\substack{j=1 \\ j\neq i}}^{N}(x_j - x_j')^2]^{1/2}$$

$$r_i'' = [\sum_{\substack{j=1 \\ j\neq i}}^{N}(x_j - x_j'')^2]^{1/2}$$

Denote the plane constructed by x_i, r_i as $\boldsymbol{P}x_i,r_i$. Then the mapping from space $\mathbf{R}^N$ to $\boldsymbol{P}x_i,r_i$ is called an Sp mapping that is formed by x_i. It is denoted as

$$\text{Sp: } \mathbf{R}^N \rightarrow \boldsymbol{P}x_i,r_i \quad (i = 1, 2, ..., N) \tag{14}$$

It has the following characteristics.

Characteristic 1. The hypersphere $F = 0$ in $\mathbf{R}^N$ degenerated as a circle $f_i = 0$ on plane $\boldsymbol{P}x_i,r_i$, here

$$f_i = (r_i - r_i^0)^2 + (x_i - x_i^0)^2 - R^2 \tag{15}$$

That is to say, if two imaged data points on plane $\boldsymbol{P}x_i,r_i$ are separated by circle $f_i = 0$, then their inverse-image point in space $\mathbf{R}^N$ should also be separated by hypersphere $F = 0$.

Characteristic 2. The imaged point (x_i^*, y_i^*) on plane $\boldsymbol{P}x_i,r_i$ corresponds with its inverse-image point $(x_1, x_2, ..., x_N)$ in space $\mathbf{R}^N$ one-to-one; the relationship between them is

$$x_k = \begin{cases} \dfrac{(r_i^* - r_i^0)(x_k' - x_k^0)}{[\sum_{\substack{j=1 \\ j\neq i}}^{N}(x_j' - x_j^0)^2]^{1/2}} + x_k^0 & (k \neq i) \\ x_i^* & (k = i) \end{cases} \qquad (k = 1, 2, ..., N) \tag{16}$$

(2) Related Derivation and Discussion. The formula 15 is but the reform of formulas 4 and 13. Characteristic 2 gives the inverse mapping from $\boldsymbol{P}x_i,r_i$ to $\mathbf{R}^N$. Here gives the derivation.

On the plane determined by axis x_i and the reference point C', the parametric function of axis r_i is

$$\frac{x_j - x_j'}{x_j - (2x_j^0 - x_j')} = t \quad (j = 1, 2, ..., N, j \neq i) \tag{17}$$

which gives

$$x_j = \frac{2tx_j^0 - (t+1)x_j'}{t-1} \quad (j \neq i) \tag{18}$$

Derive from (13)

$$(r_i - r_i^0)^2 = \sum_{j\neq i}(x_j - x_j^0)^2 \tag{19}$$

Substitute x_j from (18) and let

$$[\sum_{j\neq i}(x_j^0 - x_j')^2]^{1/2} = d_i^0 \tag{20}$$

Then formula (19) can be transformed as

$$\left(\frac{t+1}{t-1}\right)^2 = \left(\frac{r_i - r_i^0}{d_i^0}\right)^2 \tag{21}$$

Because $C^\circ \neq C' \neq C''$ (definition), from formula (17) we have $t \neq \pm 1$.

1. When $-1 < t < 1$, from (17) we have

$$\sum_{j\neq i}(x_j - x_j')^2 < \sum_{j\neq i}[x_j - (2x_j^0 - x_j')]^2$$

Therefore, from (13)

$$r_i - r_i^0 = [\sum_{j\neq i}(x_j - x_j^0)^2]^{1/2} > 0$$

Formula 21 can be transformed as

$$\frac{t+1}{1-t} = \frac{r_i - r_i^0}{d_i^0} \tag{21a}$$

II. When $t > 1$ or $t < -1$, from (17) we have

$$\sum_{j\neq i}(x_j - x_j)^2 > \sum_{j\neq i}[x_j - (2x_j^0 - x_j)]^2$$

Therefore, from (13)

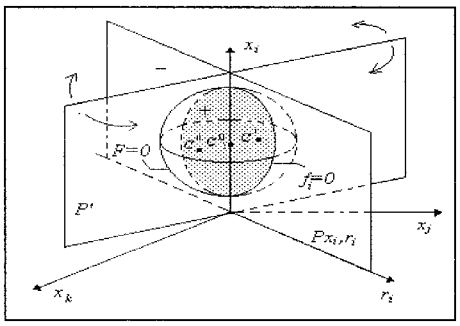

Figure 7. The geometrical meaning of Fp mapping.

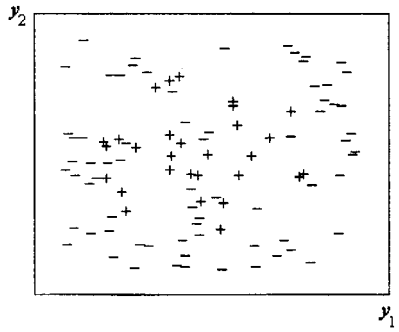

Figure 8. A PCA projection map of the testing data set. $y_1 = 0.4659x_1 + 0.4066x_2 + 0.7859x_3$. $y_2 = 0.7848x_1 + 0.2204x_2 - 0.5792x_3$.

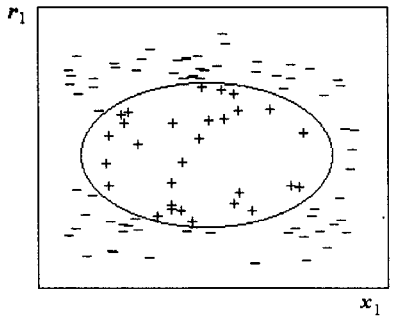

Figure 9. A Sp map of the testing data set. $r_1 = \{[(x_2 - 0.5)^2 + (x_3 - 0.5)^2]^{1/2} + 0.707\ (r' \le r'')$; $-[(x_2 - 0.5)^2 + (x_3 - 0.5)^2]^{1/2} + 0.707\ (r' > r'')\}$, $C' = (0,0,0)$, $C'' = (1,1,1)$; $r' = x_2^2 + x_3^2$, $r'' = (x_2 - 1)^2 + (x_3 - 1)^2$.

$$r_i - r_i^0 = -[\sum_{j \neq i}(x_j - x_j^0)^2]^{1/2} < 0$$

Formula 21 can be transformed as

$$\frac{t+1}{t-1} = \frac{r_i^0 - r_i}{d_i^0} \qquad (21b)$$

Formulas 21a and 21b are identical. So that

$$t = \frac{r_i - r_i^0 - d_i^0}{r_i - r_i^0 + d_i^0} \qquad (22)$$

Substitute (22) for (18) and replace the subscript j with k, we obtain

$$x_k = \frac{(r_i^* - r_i^0)(x_k' - x_k^0)}{d_i^0} + x_k^0 \quad (k \neq i) \qquad (23)$$

The derivation of characteristic 2 is done.

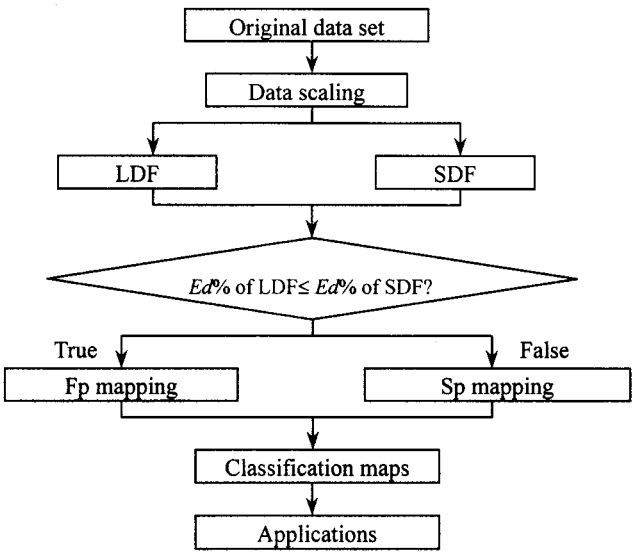

Figure 10. The computerized process of the construction of CMs (simplified).

(3) The Geometrical Meaning of Sp Mapping. See Figure 7. Space $\mathbf{R}^N$ is mapped onto plane $\boldsymbol{P}x_i,r_i$ by processing a revolving transformation with the symmetrical axis x_i. Verged by plane P' (which is the vertical plane of $\boldsymbol{P}x_i,r_i$), samples in the half space that include point C' are imaged onto the side $r_i \ge 0$ of $\boldsymbol{P}x_i,r_i$; samples in the half space that include point C'' are imaged onto the side $r_i < 0$ of $\boldsymbol{P}x_i,r_i$. Thus the hypersphere $F = 0$ degenerated as the circle $f_i = 0$ on plane $\boldsymbol{P}x_i,r_i$. As mentioned above, Sp mapping is suitable for the data set with included type classification structure, while a linear mapping such as PCA is incapable for this case. A three-dimensional testing data set with 100 samples was generated to show this case. All samples were generated randomly with limiting $x_j \in [0,1]$ ($j = 1, 2, 3$). It was classified beforehand by the following spherical function:

$$F = (x_1 - 0.5)^2 + (x_2 - 0.5)^2 + (x_3 - 0.5)^2 - 0.8^2$$

In the case of $F(P_i) < 0$, P_i is denoted as "+", otherwise it is denoted as "−". Thus, we got a typical included type data set. Figure 8 shows result mapping given by PCA, at which the data points are mixed together. Figure 9 shows a classification map constructed by Sp mapping, at which the two classes are separated completely.

3. COMPUTERIZED PROCESS

Figure 10 shows the computerized process of CMs construction. When a data set is given and is classified properly, we can first determine its classification structure by calculating its Ed% of LDF and SDF (see formula 2). If a data set has an Ed% of LDF not more than that of SDF, then it is more likely to have a partitioned type classification structure and can be processed with the Fp mapping to construct its CMs. Otherwise, the Sp mapping should be processed. In case both Ed% are greater than a preset threshold value (e.g. 20%), we should assume that the classification structure of the given data set is neither a partitioned type nor an included type, and the analytic process should be terminated. However, such a case does not occur frequently in practice. Figure 11 shows the process of

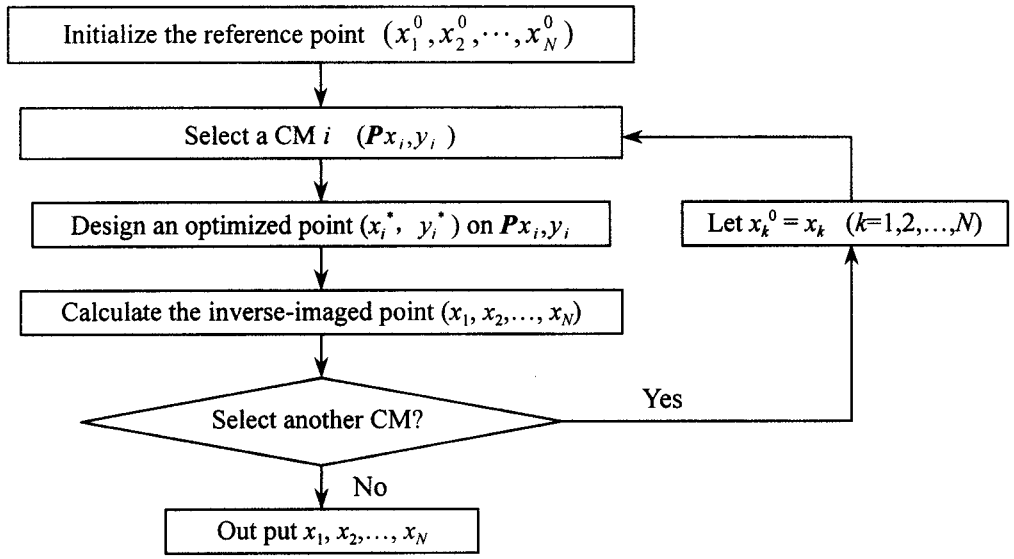

Figure 11. The computerized process of designing an optimized point on CMs that given by Fp mapping.

designing an optimized point on CMs that given by Fp mapping. The process can also be fit to Sp mapping by changing the reference point from $(x_1^0, x_2^0, \ldots, x_N^0)$ to $(x_1', x_2', \ldots, x_N')$.

4. APPLICATIONS

4.1. The Application of Fp Mapping in the Study of LSHE Propellant. 4.1.1. Problem. The characteristic σ of rocket propellant LSHE was studied experimentally. It is composed of three principal components *A*, *B*, and *C* (>15% in molar), a small quantity of additive *D* (<1.1% in molar), and other unconsidered microelements. Another factor also related with σ is the technical parameter *T* (0.1−0.2). The maximum experimental value of σ is 162 up to now (see Table 1), while in the project it is expected to reach a level of 165. The classification mapping is applied to study relationships between σ and its affected features based on what can we seek for further optimization of the material.

4.1.2. Analytical Process and Results. The analytical process can be summarized as follows.

Step 1. Data classifying: Classify samples with $\sigma > 157$ to Ω_+ and samples with $\sigma \leq 157$ to Ω_-.

Step 2. Data scaling: Let

$$x_i^{\text{new}} = \frac{x_i^{\text{old}} - \bar{x}_i}{S_i} \quad (i = 1, 2, \ldots, 5)$$

x_1, x_2, x_3, x_4, and x_5 correspond with the percentage of *A*, *B*, *C*, and *D*, and the value of *T*, respectively. $\bar{x}_i$ is the mean value of x_i and S_i is the standard deviation of x_i. The scaling process eliminated the difference in the order of magnitude of these features.

Step 3. LDF construction: Work out LDF of the data set by formula 3

$$F = 0.44A - 0.28B - 0.30C - 0.78D - 36.07T - 2.93$$

The classification error of this discriminant function Ed% = 10%, only three samples are misclassified.

Step 4. Classification maps construction: Calculate the y_i for each x_i by formula 5 and mark all samples to plane constructed by x_i, y_i, thus we get five classification maps (Figure 12). The samples in different classes are separated clearly on these maps. Figure 13 shows one of them. Figure 14 is a PCA map constructed by the first two primary PCs. Affected by nonlinear relationships that existed between σ and its controlling factors, the PCA maps can not show the classification distinctly.

Table 1. Experimental Results of LSHE Studies

		features					target
no.	class	*A* (%)	*B* (%)	*C* (%)	*D* (%)	*T*	σ
1	2	52.0	14.5	32.5	0.84	0.10	153
2	1	48.0	21.0	30.5	0.05	0.16	158
3	2	53.0	16.0	30.0	0.43	0.10	151
4	2	52.5	14.7	31.5	0.93	0.10	153
5	1	48.0	18.5	32.5	0.82	0.12	159
6	1	46.0	19.0	33.5	0.83	0.16	162
7	1	48.5	18.0	32.5	0.76	0.16	161
8	2	55.0	15.5	29.0	0.42	0.15	154
9	1	51.5	17.5	30.5	0.84	0.15	159
10	1	49.0	16.0	34.0	0.59	0.14	157
11	1	49.5	15.5	34.5	0.51	0.18	160
12	2	48.5	17.0	34.0	0.24	0.10	153
13	2	51.0	20.5	27.5	0.65	0.10	156
14	2	51.5	15.5	32.0	0.30	0.14	153
15	1	48.5	21.0	30.0	0.09	0.16	158
16	2	50.0	16.0	33.5	0.09	0.15	154
17	1	46.0	19.0	34.5	0.64	0.10	158
18	2	50.5	15.5	33.0	0.81	0.12	155
19	1	48.0	21.5	30.0	0.59	0.10	158
20	2	45.5	22.0	32.0	0.01	0.12	155
21	2	50.5	17.5	31.0	0.72	0.11	156
22	2	53.5	15.0	30.5	0.43	0.18	156
23	2	48.0	18.5	33.5	0.05	0.12	153
24	2	52.5	17.5	28.5	0.84	0.12	156
25	1	51.0	17.0	31.0	0.69	0.17	160
26	2	54.0	14.5	30.0	0.73	0.10	151
27	1	48.5	21.0	30.0	0.06	0.17	158
28	1	50.0	14.0	35.5	0.55	0.17	157
29	2	52.5	18.5	29.0	0.10	0.13	152
30	2	52.0	18.0	29.5	0.42	0.11	154

Figures 12 and 13 show remarkable regularity. It can be applied in the following works.

(1) Simulation of an experiment: Design an experimental scheme and mark it into CMs. If it is located in a "bad" region, then the scheme can be rejected. For example, an experimental scheme is designed as $A = 53.0\%$, $B = 16.0\%$, $C = 29.5\%$, 0.27%, and $T = 0.13$, the imaged point is calculated by formula 5 and marked as point "?" on Figure

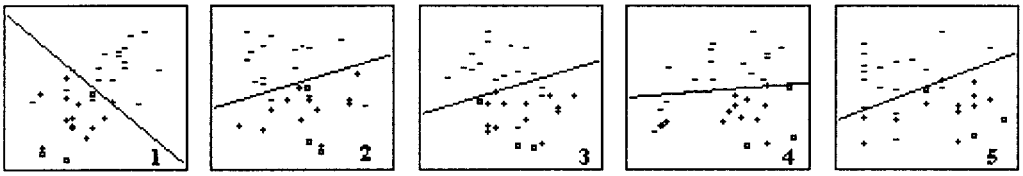

Figure 12. The five classification maps of the HELS data given by Fp mapping: (+) $\sigma > 157$, (−) $\sigma \leq 157$, and (□) the three designed schemes.

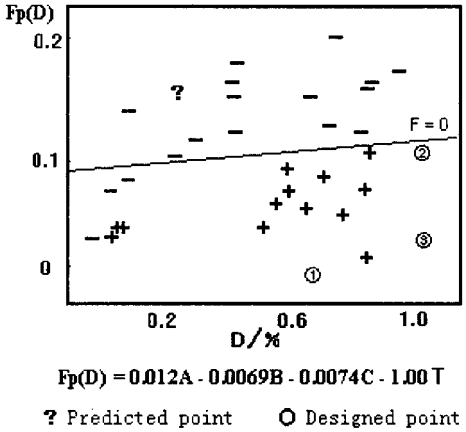

Figure 13. A classification map constructed by Fp mapping: (+) $\sigma > 157$ and (−) $\sigma \leq 157$.

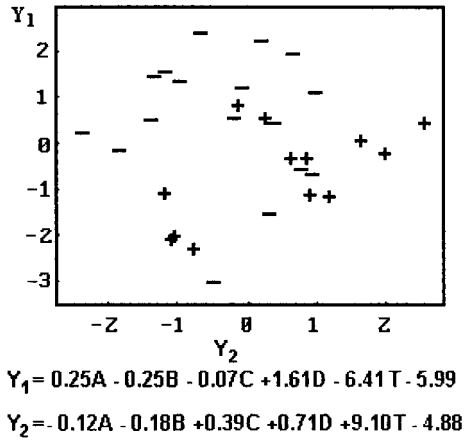

Figure 14. A projection map constructed by PCA mapping: (+) $\sigma > 157$ and (−) $\sigma \leq 157$.

13. According to characteristic 1 of Fp mapping, we know that its inverse-image point should be located in the Ω_- region and likely to have a σ value less than 157, so that the scheme can be abandoned.

(2) Design of an experiment: Design optimized points on classification maps and calculate their inverse-image point according to characteristic 2 of Fp mapping, from which we can get optimized experimental schemes. For example, by spotting a point "circled one" on the "good" region of Figure 13 and calculating its corresponding experimental parameters by formula 9, we have $A = 46.0\%$, $B = 19.0\%$, $C = 33.0\%$, $D = 0.68\%$, and $T = 0.17$. The σ value of sample prepared according to this scheme is measured as 163.2. Two more schemes are designed on the CMs (point "circled two-","circled three" on Figure 13), and the experimental results are listed in Table 2. A significant improvement has been achieved, and the expected target of the project ($\sigma \geq 165$) has been reached.

Table 2. Designed Schemes and the Experimental σ of LSHE Material

no.	*A* %	*B* %	*C* %	*D* %	*T*	exptl σ
1	46.0	19.0	33.0	0.68	0.17	**163.2**
2	50.0	18.0	29.5	1.00	0.14	**165.3**
3	48.0	18.5	32.0	1.02	0.19	**166.5**

Table 3. Experimental Results of Weld Material

no.	Mn	Si	S	Ti	B	N	O	Sy (N/mm^2)	St (N/mm^2)	Tr (°C)
1	1.50	0.29	0.008	0.012	0.0002	0.016	0.034	510	570	−29
2	1.53	0.27	0.005	0.012	0.0032	0.017	0.032	530	580	−36
3	1.53	0.28	0.005	0.013	0.0057	0.017	0.031	520	560	−37
4	1.47	0.27	0.005	0.011	0.0090	0.015	0.027	480	550	−43
5	1.43	0.29	0.005	0.011	0.0129	0.015	0.031	470	530	−46
6	1.45	0.35	0.005	0.010	0.0158	0.017	0.031	450	520	−47
7	1.51	0.39	0.007	0.030	0.0002	0.017	0.028	550	610	−35
8	1.54	0.37	0.004	0.020	0.0046	0.016	0.029	510	570	−36
9	1.51	0.42	0.005	0.021	0.0105	0.016	0.027	500	560	−43
10	1.61	0.39	0.004	0.021	0.0147	0.017	0.027	500	550	−48
11	1.50	0.41	0.004	0.023	0.0200	0.014	0.027	490	570	−50
12	1.48	0.47	0.007	0.041	0.0002	0.016	0.029	580	630	−44
13	1.53	0.37	0.006	0.037	0.0015	0.015	0.029	550	610	−48
14	1.51	0.38	0.007	0.041	0.0029	0.014	0.029	530	590	−48
15	1.48	0.34	0.007	0.037	0.0040	0.015	0.028	520	580	−48
16	1.46	0.34	0.006	0.035	0.0054	0.014	0.028	520	580	−52
17	1.48	0.35	0.007	0.035	0.0102	0.014	0.028	500	560	−64
18	1.39	0.35	0.006	0.036	0.0158	0.014	0.029	490	560	−67
19	1.45	0.28	0.009	0.012	0.0002	0.024	0.032	520	600	−23
20	1.46	0.35	0.005	0.012	0.0029	0.024	0.031	530	590	−26
21	1.45	0.34	0.005	0.012	0.0059	0.025	0.034	540	590	−22
22	1.39	0.33	0.005	0.011	0.0090	0.027	0.031	520	580	−29
23	1.47	0.30	0.004	0.011	0.0128	0.024	0.032	500	550	−34
24	1.37	0.30	0.005	0.010	0.0162	0.021	0.031	470	530	−35
25	1.48	0.40	0.006	0.032	0.0002	0.025	0.029	550	620	−24
26	1.47	0.32	0.005	0.019	0.0043	0.027	0.025	530	590	−18
27	1.41	0.34	0.005	0.021	0.0091	0.024	0.026	510	580	−22
28	1.44	0.34	0.005	0.021	0.0143	0.024	0.027	490	560	−34
29	1.37	0.40	0.006	0.025	0.0200	0.022	0.027	500	550	−46
30	1.46	0.40	0.007	0.043	0.0002	0.023	0.030	550	630	−30
31	1.51	0.38	0.007	0.041	0.0016	0.023	0.029	540	630	−30
32	1.42	0.36	0.007	0.038	0.0028	0.022	0.030	540	610	−27
33	1.51	0.36	0.007	0.041	0.0044	0.023	0.029	560	620	−24
34	1.46	0.36	0.007	0.037	0.0052	0.023	0.038	540	600	−25
35	1.44	0.35	0.007	0.036	0.0110	0.023	0.029	490	560	−38
36	1.40	0.36	0.007	0.039	0.0167	0.022	0.030	470	540	−52

4.2. The Application of Sp Mapping in the Study of a Kind of Welding Material. 4.2.1. Problem. The relationship between the chemical composition and the mechanical properties of a kind of welding material was studied experimentally. The study objects are yield strength of the welded joint (Sy), tensile strength of the welded joint (St), and the transformation temperature of KCV 100J percussive work (Tr). The affecting features are seven chemical components: Si, Mn, S, Ti, B, N, and O. The amounts of other elements are controlled nearly constant. The original data set contains 36 samples (see Table 3). The target of this project is to seek a suitable composition to enable the material's mechanical properties to reach the following values: St > 580 N/mm^2 and Sy > 500 N/mm^2 and Tr <

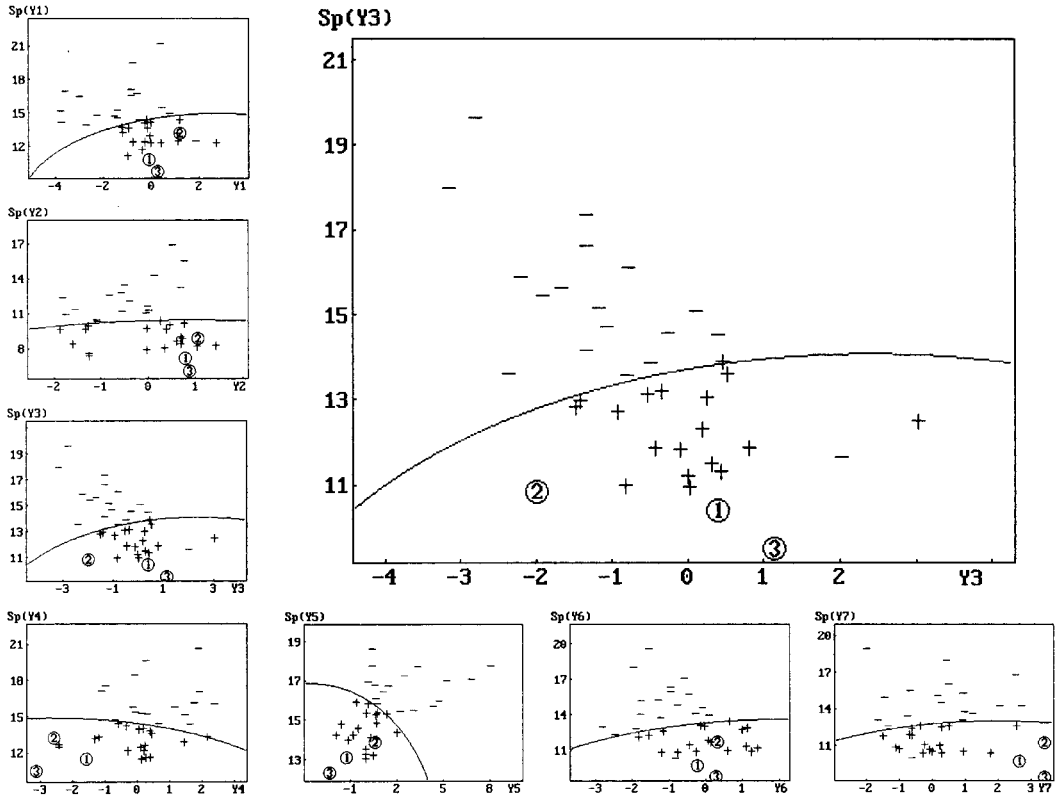

Figure 15. The seven classification maps of the welding material data given by Sp mapping: (−) "bad" samples (Ω_-), (+) "good" samples (Ω_+), and (○) designed samples.

−60 °C. Sp mapping is applied here to solve the problem.

4.2.2. Analytical Process and Results. Step 1. Data classifying: Define a summarized index first

$$T = Q\left(\frac{\mathrm{St} - 580}{\mathrm{Max(St)} - \mathrm{Min(St)}}\right) + Q\left(\frac{\mathrm{Sy} - 500}{\mathrm{Max(Sy)} - \mathrm{Min(Sy)}}\right) + kQ\left(\frac{-60 - \mathrm{Tr}}{\mathrm{Max(Tr)} - \mathrm{Min(Tr)}}\right)$$

The coefficient $k = 10$, here it is set to emphasize the essential property Tr. Function Q is defined as

$$Q(X) = \begin{cases} |X|^{1/2} & \text{when } X \geq 0 \\ -|X|^{1/2} & \text{when } X < 0 \end{cases}$$

Divide the given samples into two classes average based on the calculated index T and classify samples with $T > -0.365$ to Ω_+ and samples with $T \leq -0.365$ to Ω_-.

Step 2. Data scaling: Let

$$x_i^{\mathrm{new}} = \frac{x_i^{\mathrm{old}} - \bar{x}_i}{S_i} \quad (i = 1, 2, ..., 7)$$

x_1, x_2, x_3, x_4, x_5, x_6, x_7 correspond with the percentage of Si, Mn, S, Ti, B, N, and O, respectively. $\bar{x}_i$ is the mean value of x_i, and S_i is the standard deviation of x_i.

Step 3. SDF construction: The whitening transformation and genetic algorithms are applied to construct the SDF (see section 2.2.2)

$$F = \sum_{j=1}^{7} (Y_j)^2 - 67.142 < 0$$

Here

$$Y_1 = +14.13\,\mathrm{Mn} - 0.101\,\mathrm{Si} + 34.06\,\mathrm{S} + 32.94\,\mathrm{Ti} - 11.61\,\mathrm{B} - 142.2\,\mathrm{N} + 416.0\,\mathrm{O} - 31.39$$

$$Y_2 = +1.402\,\mathrm{Mn} + 6.409\,\mathrm{Si} + 71.53\,\mathrm{S} + 30.51\,\mathrm{Ti} + 1.143\,\mathrm{B} - 71.97\,\mathrm{N} - 146.4\,\mathrm{O} + 0.202$$

$$Y_3 = +0.064\,\mathrm{Mn} - 13.02\,\mathrm{Si} + 875.9\,\mathrm{S} + 6.143\,\mathrm{Ti} - 24.02\,\mathrm{B} - 8.894\,\mathrm{N} - 42.50\,\mathrm{O} + 0.615$$

$$Y_4 = +12.00\,\mathrm{Mn} - 21.22\,\mathrm{Si} - 432.9\,\mathrm{S} + 104.5\,\mathrm{Ti} + 96.99\,\mathrm{B} + 208.9\,\mathrm{N} - 119.0\,\mathrm{O} - 12.46$$

$$Y_5 = +29.94\,\mathrm{Mn} + 24.19\,\mathrm{Si} + 833.5\,\mathrm{S} - 96.38\,\mathrm{Ti} + 409.4\,\mathrm{B} + 135.5\,\mathrm{N} - 64.45\,\mathrm{O} - 55.32$$

$$Y_6 = +0.063\,\mathrm{Mn} + 10.52\,\mathrm{Si} + 197.8\,\mathrm{S} + 21.78\,\mathrm{Ti} - 50.91\,\mathrm{B} + 130.2\,\mathrm{N} + 80.90\,\mathrm{O} - 9.960$$

$$Y_7 = -15.44\,\mathrm{Mn} + 1.626\,\mathrm{Si} + 112.6\,\mathrm{S} + 31.78\,\mathrm{Ti} + 101.3\,\mathrm{B} - 20.87\,\mathrm{N} + 155.7\,\mathrm{O} + 15.52$$

The classification error of this discriminant function Ed% = 5.5%, only two samples are misclassified (see Figure 15). Or we can say approximately, a supposed material with calculated $F < 0$ should have a better performance than that of calculated $F > 0$.

Step 5. Classification maps construction: Calculate the r_i for each x_i by formula 13 and mark all samples to plane constructed by x_i, r_i, thus we get seven classification maps (Figure 15).

The coordinate axeses in Figure 15 are

$$\begin{cases} \mathrm{Sp}(Y_3) = r + r_0 & \text{when } r' \le r'' \\ \mathrm{Sp}(Y_3) = -r + r_0 & \text{when } r' > r'' \end{cases}$$

here

$$r = \left(\sum_{\substack{j=1 \\ j\neq 3}}^{7} (Y_j - X_j^0)^2\right)^{1/2} \quad r_0 = \left(\sum_{\substack{j=1 \\ j\neq 3}}^{7} (X_j^0)^2\right)^{1/2} = 5.88$$

$$r' = \left(\sum_{\substack{j=1 \\ j\neq 3}}^{7} (Y_j - X_j')^2\right)^{1/2} \quad r'' = \left(\sum_{\substack{j=1 \\ j\neq 3}}^{7} (Y_j - X_j'')^2\right)^{1/2}$$

For the expression of Y_j (j = 1, 2, ..., 7) see the above paragraphs; for X_j^0, X_j', and X_j'' see Table 4. The samples in different classes are separated clearly on these maps. Three points are designed artificially on the maps (see Figure 15), and the calculated compositions and experimental results of the materials are listed in Table 5. The required performances of the material have been obtained.

CONCLUSIONS

The conception and theoretical groundwork of the classification mapping have been discussed. Two basic classification mappings (Fp mapping and Sp mapping) have been proposed. By these methods, we can construct classification maps from multivariate data space, describe the classification structure of data set intuitively, and keep a strict relationship with the original data space. They are not only display methods but also calculation and extrapolation tools. The CMs may be regarded as a kind of geometric model assisting in multivariate analysis. It can be applied in fields such as multivariate data processing in chemical analysis, componential and technical exploration for materials, or experimental and productional optimization. These methods were applied to the optimization of several materials and achieved significant improvement on their performances.

Table 4. The Value of Parameters X_j^0, X_j', and X_j''

j	X_j^0	X_j'	X_j''
1	2.75310	−1.31251	6.81872
2	0.92268	−0.51790	2.36327
3	2.39921	−1.11900	5.91742
4	−2.73627	0.46007	−5.93261
5	−3.53321	2.52106	−9.58748
6	2.20897	−1.19431	5.61225
7	1.96441	0.23323	3.69558

Table 5. Designed Schemes and the Experimental Result of the Welding Material

no.	Mn	Si	S	Ti	B	N	O	Sy (N/mm²)	St (N/mm²)	Tr (°C)	calcd SDF
1	1.38	0.39	0.0074	0.043	0.0086	0.014	0.030	510	578	−60	−42.32
2	1.39	0.47	0.0061	0.046	0.0121	0.013	0.033	517	583	−61	−22.99
3	1.35	0.40	0.0085	0.046	0.0065	0.013	0.032	514	588	−62	−52.00

REFERENCES AND NOTES

(1) Cremperline, P. J.; Webber, L. D.; et al. *Anal. Chem.* **1989**, *61*, 138.
(2) Chen, N. Y.; et al. *J. Alloys Compds.* **1996**, *234*, 125.
(3) Colender, U. E.; et al. *Logical and Combinatorial Algorithms for Drug Design*; London, 1985.
(4) Otto, M. *Mikrochim. Acta [wien]* **1987**, *1*, 445.
(5) Huber, P. J. *Annals Statistics* **1985**, *13*, 433.
(6) Jolliffe, T. *Principal Component Analysis*; Spinger Verlag: New York, 1986.
(7) Liu, H. L.; Chen, Y.; et al. *J. Chemometrics* **1994**, *8*, 439.
(8) Strouf, O. *Chemical Pattern Recognition*; Research Studies: Letchoworth, 1986.
(9) Wang, B. Q.; Wang, C. Z. *Acta Seismologica Sinica [Chinese]* **1988**, *10*, 113.
(10) Fukunaga, K. *Introduction to Statistical Pattern Recognition*, 2nd ed.; Academic Press, Inc.: London, 1990; Chapters 2 and 5.
(11) Khots, M. S., Bazarov, V. I.; Emelyanov, V. E. *J. Chemometrics* **1995**, *9*, 411.
(12) Holland, J. H. *Adaptation in Nature and Artificial Systems*; The University of Michigan Press: Michigan, 1975.
(13) Michalewicz, Z.; et al. *Computer and Mathematics with Applications*; **1992**, *23*, 83.

CI990302O