

## 패턴인지법에 의한 한국산 고대 유리제품의 분류

李 澈\* · 蔡命俊 · 金昇源 · 姜炳台<sup>†</sup> · 李鍾斗<sup>‡</sup>

한양대학교 자연과학대학 화학과

<sup>†</sup> 국립문화재연구소

<sup>‡</sup> 한국원자력연구소

(1991. 8. 5 접수)

## Classification of Korean Ancient Glass Pieces by Pattern Recognition Method

Chul Lee\*, Myung-Zoon Czae, Seungwon Kim, Hyung Tae Kang<sup>†</sup>, and Jong Du Lee<sup>‡</sup>

Department of Chemistry, Hanyang University, Seoul 133-791, Korea

<sup>†</sup>National Research Institute of Cultural Properties, Seoul 110-050, Korea

<sup>‡</sup>Korean Atomic Energy Research Institute, Seoul 139-240, Korea

(Received August 5, 1991)

**요 약.** Chemometrics의 한 분야인 패턴인지(pattern recognition)법을 한국산 고대 유리시료 94종의 중성자방사화분석으로부터 얻은 다변수데이터에 적용하였다. unsupervised learning의 방법인 주성분분석법과 비선형도시법으로 시료를 분류한 결과 유리시료는 6개의 군을 형성하였다. 6개의 참조시료셋트와 시험시료셋트에 supervised learning의 SIMCA법을 적용시켰다. 그 결과 참조시료셋트는 주성분분석법 및 비선형도시법의 결과와 일치하였고 시험시료셋트에서 33개의 시료 중 17개 시료에 대해 시료가 속한 군을 판정할 수 있었다.

**ABSTRACT.** The pattern recognition methods of chemometrics have been applied to multivariate data, for which ninety four Korean ancient glass pieces have been determined for 12 elements by neutron activation analysis. For the purpose, principal component analysis and non-linear mapping have been used as the unsupervised learning methods. As the result, the glass samples have been classified into 6 classes. The SIMCA (statistical isolinear multiple component analysis), adopted as a supervised learning method, has been applied to the 6 training set and the test set. The results of the 6 training set were in accord with the results by principal component analysis and non-linear mapping. For test set, 17 of 33 samples were each allocated to one of the 6 training set.

### 서 론

최근 분석 기술의 발달로 소량의 시료로부터 많은 양의 분석데이터를 얻을 수 있으며 데이터의 해석을 위하여 Chemometrics<sup>1,2</sup>가 많이 응용되고 있다. Chemometrics는 데이터를 해석하고 결과를 예측하는 통계적인 방법이다.

Chemometrics 방법 중에 패턴인지법(pattern recognition)<sup>3</sup>은 시료에서 얻은 정성적, 정량적 정보

로부터 직접 측정은 불가능하지만 미지의 상호관계에 의해 관련된 시료의 성질을 발견하는 방법이다. 이 방법은 고고학적 시료에 적용<sup>4-6</sup> 될 뿐 아니라 환경과학<sup>7</sup>, 범죄과학<sup>8</sup> 및 IR 스펙트럼의 해석<sup>9,10</sup> 등에 널리 응용되고 있다.

본 연구에서는 전보<sup>11</sup>에서 분석된 42종의 유리시료와 새로이 입수된 52종의 시료로부터 중성자방사화분석법에 의해 12종의 성분 원소의 함량을 결정

하였다. 이들 분석데이터는 unsupervised learning 를 위해 전보에서 사용한 주성분분석법(principal component analysis)<sup>6,8</sup> 뿐 아니라 본 저자에 의해 새로이 개발한 비선형도시법(non-linear mapping)<sup>4,12</sup> 을 사용하여 전체 94종의 시료가 몇 개의 참조시료셋으로 분류되는지 먼저 확인한 다음 전보의 결과와 비교 검토하였다. 참조시료셋에 속하지 않는 시료는 시험시료셋으로 취급하였으며 SIMCA(statistical isolinear multiple component analysis)<sup>2,13</sup> 법을 사용하여 각 셋트의 시료가 통계적으로 기대되는 신뢰성으로 분류되는지 종합적으로 검토하였다.

## 실 험

입수한 유리 시료는 전보<sup>11</sup>와 동일한 방법으로 전처리하였다. 단일비교체로써 함량을 알고 있는 Au 및 Co를 부착하여 한국원자력연구소의 TRIGAMARK III 원자로의 회전 시료 조사대에서 15시간 중성자로 조사하였다. 이 조사위치에서의 중성자속은  $1 \times 10^{13}$  neutrons  $\text{cm}^{-2}\text{sec}^{-1}$ 이었다.

감마선 계측은 4000 채널 다중과고분석기(multi-channel analyzer, ORTEC Model 7050)와 전자계산조직에 연결되어 있는 75 cc 용적의 Ge(Li) 반도체 검출기를 사용하였다. 계측된 감마선의 에너지와 피크면적을 계산하였고 기하학적 위치에 따른 계측 효율, 해석적인 데이터 및 조사조건에 따른 중성자속을 이용하여 각 성분 원소의 함량을 단일비교체법으로 결정하였다<sup>14,15</sup>.

상기 중성자방사화분석법은 미량원소의 동시 분석이 가능하고 시료가 비파괴적이라는 장점이 있어 널리 쓰이고 있다. 또한 대부분의 원소의 정량치가  $\pm 10\%$  이내로 그 정확성 및 정밀성이 확인된 바 있다. 본 연구에서는 각 시료를 두번 분석한 다음 그 평균치를 각 시료에 포함된 원소의 함량으로 결정하였다.

전보에서는 20개 원소의 성분함량을 결정하였으나 본 연구에서는 12개 원소의 성분 함량을 결정하였다. 이는 유리의 주성분인 K 및 Na는 방해물 없애기 위하여 조사 후 냉각에 의해 제거되었고 주성분분석법의 결과 공동성(communality)에 의한 기여도가 작은 것으로 확인된 As, Ru이 또한 제외되었을 뿐 아니라 본 연구에서 측정하지 못한 Pb, Sb, Zn 및

Ag가 제외되었기 때문이다.

## 패턴인지법(Pattern Recognition Method)의 기초이론

**주성분분석법 및 비선형도시법.**  $n$ 개의 시료에서 분석된  $m$ 개의 변수  $x_{ij}$ 는 한 셋트의 매트릭스  $X_{ij}$  ( $n \times m$ )를 구성한다. 이는  $m$  차원의 공간에  $n$ 개의 시료점을 나타낸 것과 같다.  $m$  차원에 나타나 있는 시료의 분포형태를 육안으로 인지할 수 있도록 2, 3차원으로 나타내는 것이 필요하다. 만약 시료에 대한 사전 정보가 없는 경우에 이를 unsupervised learning 방법이라하며 주성분분석법과 비선형도시법이 있다. 이의 결과로 각 시료들간의 관계를 알 수 있으며 시료가 몇 개의 군으로 분류되어짐을 알 수 있다.

주성분분석법의 경우 측정된 모든 변수가 동등한 중요성을 가지도록 원변수  $x_{ij}$ 을 새로운 변수  $z_{ij}$ 로 변환하는 것이 필요하다. 이런 과정을 autoscaling이라 하며 식 (1)과 같이 나타낸다.

$$z_{ij} = \frac{x_{ij} - x_j}{\sigma_j} \quad (1)$$

여기서  $i$ 는 1, 2, 3, ...,  $n$ 이며 시료의 수를 나타내고,  $j$ 는 1, 2, 3, ...,  $m$ 으로 변수의 수를 나타낸다. 매트릭스 기술법으로 표시하면  $[Z] = (z_{ij})_{n \times m}$ 이 된다. 여기서  $x_{ij}$ 는 각 시료  $i$ 에서 측정된 변수  $j$ 의 성분 벡터를 나타내며,  $x_j$ 와  $\sigma_j$ 는 변수  $j$ 의 전체평균과 표준편차를 나타낸다.  $n$ 개의 시료에 대한  $m$ 개의 변수로부터 얻을 수 있는 주성분값은 식 (2)와 같이 선형결합으로 구할 수 있다.

$$y_{ij} = k_1 z_{i1} + \dots + k_j z_{ij} + \dots + k_m z_{im} \quad (2)$$

이를 매트릭스 기술법으로 표시하면  $[Y] = [K][Z]$ 가 되면  $[K]$ 는 새로운 변수의 고유벡터의 계수 매트릭스이다.  $[Z]$ 는 바뀌어진 새로운 변수  $z_{ij}$ 의 데이터 매트릭스이고  $[Y]$ 는 주성분값을 나타내는 매트릭스이다.

식 (2)의 고유벡터의 계수 매트릭스는 다음과 같이 구할 수 있다. 새로운 변수  $z_{ij}$ 의 데이터 매트릭스  $[Z]$ 와 이를 전환시킨 전환 매트릭스  $[Z]^T$ 로부터 분산-공분산 매트릭스  $[R]$ 을 식 (3)과 같이 구할 수 있다.

$$[R]=[Z]^T[Z] \quad (3)$$

매트릭스  $[R]$ 로부터 식 (4)의 조건을 만족시키면 매트릭스  $[R]$ 를 대각선회할 수 있다.

$$[K]^{-1}[R][K]=\lambda_j\delta_{jk} \quad (4)$$

$\lambda_j$ : eigenvalue,  $\lambda_1 > \lambda_2 > \dots > \lambda_m$

$\delta_{jk}$ : Kronecker delta =  $\begin{cases} 0 & \text{if } j \neq k \\ 1 & \text{if } j = k \end{cases}$

식 (4)를 다시 쓰면 식 (5)와 같이 된다.

$$[R][K]=\lambda_j[K] \quad (5)$$

고유값  $\lambda_j$ 에 대응하는 제수벡터  $K_j$ 는 서로 직교하며  $\lambda_j$ 의 크기에 따라 순차적으로 구할 수 있다.

Sammon<sup>12</sup>에 의해서 처음 제안된 비선형도시법은 시료의 분포 즉, 시료간의 호트러짐을 최소화시키는 것이 중요한 문제이다.  $m$ 차원( $m$ 개의 변수)에서 시료간의 거리와 2차원(2개의 변수)에서의 시료간의 거리는 식 (6), (7)로써 계산할 수 있다.

$$d_{ik}^* = \left[ \sum_{j=1}^m (x_{ij} - x_{kj})^2 \right]^{1/2} \quad (6)$$

$$d_{ik} = \left[ \sum_{q=1}^2 (x_{iq}(l) - x_{kq}(l))^2 \right]^{1/2} \quad (7)$$

식 (7)의 경우 가장 큰 표준편차를 가진 두 개의 변수를 사용하여 시료간의 거리를 계산하는 것이 원칙이다.

비선형도시법은  $m$  차원에서의 시료간 거리와 2차원에서의 시료간 거리의 차를 나타내는 식 (8)의 mapping 오차  $E$ 를 최소화시키는 것이 필요하다. 이를 위해서 2차원의 좌표를 바꾸면서 반복계산하는 gradient method<sup>12</sup>가 식 (9)와 같이 사용된다. mapping 오차  $E$ 가 최소가 될 때의  $x_{i1}$ ,  $x_{i2}$ 를 구하여 평면에 도시하면  $m$ 차원의 분포가 2차원에 나타나게 된다.

$$E = \sum_{i < k} \frac{[d_{ik}^* - d_{ik}(l)]^2}{d_{ik}^{*2}} \quad (8)$$

$$x_{iq}(l+1) = x_{iq}(l) - MF \cdot \Delta_{iq}(l) \quad (9)$$

여기서

$$\Delta_{iq}(l) = \frac{\partial E(l)}{\partial x_{iq}(l)} / \frac{\partial^2 E(l)}{\partial x_{iq}(l)^2} \quad (10)$$

이다. MF는 magic factor로써 경험적인 값이며 대개  $MF \approx 0.3 - 0.4$ 의 값을 갖는다.

본 연구에서는 이를 개량한 방법<sup>4,16</sup>으로 식 (6)의  $x_{ij}$  및  $x_{kj}$  대신에 주성분분석법에서 계산된 주성분값  $y_{ij}$  및  $y_{kj}$ 를 이용하여 시료간의 거리를 구하였다. 가변도가 큰 주성분 1, 주성분 2를 두 축으로 사용하여 2차원에서의 시료간의 거리를 상기와 같은 방법으로 식 (7)에 의해 구하였다. 이는 두 가변도만을 사용하여 시료의 분포를 도시하는 주성분분석법의 단점을 보완하는 방법으로 전가변도를 모두 사용하여 시료의 분포를 평면에 도시한 결과가 된다.

**SIMCA법.** 분류된 시료의 각 군에 분류 규칙을 만드므로써 미지시료가 여러 군들 가운데 어느 군에 속하는지 알아낼 수가 있다. 이러한 접근법은 supervised learning이라 하며 SIMCA법이 본 연구에 사용되었다.

$n$ 개의 시료에서  $m$ 개의 변수를 측정 한 시료들이  $q$ 개의 군으로 이루어져 있음을 unsupervised learning에 의해 알았다면 각 군은 식 (11)과 같이 modeling할 수 있다.

$$y_{ik}^{(q)} = \alpha_i^{(q)} + \sum_{a=1}^{A_q} \beta_{ia}^{(q)} \theta_{ak}^{(q)} + \epsilon_{ik}^{(q)} \quad (11)$$

여기서  $y_{ik}^{(q)}$ 는  $q$ 군에 속하고 있는 시료  $k$ 속에 함유된 성분 원소  $i$ 의 autoscaling된 분석치이며,  $A_q$ 는 주성분의 수이다. 식 (11)의 모델을 사용하기 전, 즉 새 시료를 분류하기 전에 파라미터  $\alpha_i^{(q)}$ ,  $\beta_{ia}^{(q)}$ ,  $\theta_{ak}^{(q)}$ ,  $\epsilon_{ik}^{(q)}$ 를 결정하여야 한다.

$\beta_{ia}^{(q)}$  및  $\theta_{ak}^{(q)}$ 의 결정은  $Y^{(q)} Y^{(q)T}$  매트릭스에서 대각선회(diagonalization)를 이용해서 고유값 및 고유벡터를 구하는 것에 귀결된다. 이 때  $\beta_{ia}^{(q)}$ 는 변수의 분포를 나타내고  $\theta_{ak}^{(q)}$ 는 시료의 분포를 나타낸다. 잔류오차  $\epsilon_{ik}^{(q)}$ 는  $y_{ik}^{(q)}$ 값에서  $A_q$ 와 같은 수의  $\beta_{ia}^{(q)}$ 와  $\theta_{ak}^{(q)}$ 의 곱항을 빼주어 구할 수 있다. 이 잔류오차  $\epsilon_{ik}^{(q)}$ 로부터 식 (12)에 의해  $S_O^{(q)2}$ 를 구할 수 있으며, 이를 참조시료셋트  $q$ 의 가변도  $\sigma_q^2$ 로 한다.

$$S_O^{(q)2} = \sum_k \sum_i \frac{\epsilon_{ik}^{(q)2}}{(n_q - A_q - 1)(m - A_q)} \quad (12)$$

여기서  $i$ 는  $1, 2, \dots, m$ 으로  $m$ 은 분석된 원소의 수이며,  $k$ 는  $1, 2, \dots, n_q$ 이고  $n_q$ 는 참조시료셋트  $q$ 에 속한 시료의 수이다.

모델식 (11)은 각 군의 참조시료에 독립적으로 맞추어지기 전에 각 군의 차수, 즉 항의 수  $A_q$ 를 결정하여야 한다.  $\beta_{ik}^{(q)}$ 와  $\theta_{ik}^{(q)}$ 를 구한 후  $A_q$ 를 증가시키면서 각 시료에 대해 잔류오차  $\varepsilon_{ip}^{(q)}$ 의 제곱을 계산한다.  $A_q$ 를 증가시키면서 참조시료셋트  $q$ 에 속한 모든 시료의 잔류오차의 제곱을 더한  $D_A$ 를 구하여  $(D_{A-1} - D_A)/(m - A_q)$  vs.  $D_A/(n_q - A_q - 1)(m - A_q)$ 를 계산한다. 이를  $(m - A_q)$  vs.  $(n_q - A_q - 1)(m - A_q)$ 에 대응하는 임계  $F$ -값( $p=0.05$ )과 비교하므로서 항의 수  $A_q$ 를 결정한다.

어떤 시료  $p$ 의 관측치  $y_{ip}$ 를 이상에서 구한  $\alpha_i^{(q)}$ ,  $\beta_{ik}^{(q)}$  및 항  $A_q$ 를 사용하여 식 (13)에 맞추어서 잔류오차  $\varepsilon_{ip}^{(q)}$ 를 구한다.

$$y_{ip} = \alpha_i^{(q)} + \sum_{k=1}^{A_q} C_{qk} \beta_{ik}^{(q)} + \varepsilon_{ip}^{(q)} \quad (13)$$

여기서 파라미터  $C_{qk}$ 는 앞에서 구한  $\beta_{ik}^{(q)}$ 를 사용해서 잔류오차  $\varepsilon_{ip}^{(q)}$ 를 최소화시켜 줌으로써 구할 수 있다. 이와 같이 구한 잔류오차  $\varepsilon_{ip}^{(q)}$ 로부터 가변도  $S_p^{(q)2}$ 을 식 (14)로써 구할 수 있으며 시료  $p$ 가  $q$ 군에 속하는지 또는 열외인지에 관한 판단기준은  $F = S_p^{(q)2}/S_0^{(q)2}$ 에 둔다. 시료  $p$ 의 계산된  $F$ -값이  $(m - A_q)$  vs.  $(n_q - A_q - 1)(m - A_q)$ 에 대응하는 임계  $F$ -값( $p=0.05$ )보다 적으면 시료  $p$ 는  $q$ 군에 속하는 것으로 판정한다.

$$S_p^{(q)2} = \sum_{i=1}^m \frac{\varepsilon_{ip}^{(q)2}}{(m - A_q)} \quad (14)$$

#### 결과 및 고찰

시료에 대한 고고학적 데이타와 입수한 위치를 기호와 함께 Table 1에 나타내었다. 중성자방사화 분석법에 의해 측정된 12개 성분원소의 평균과 표준편차를 계산하여 Table 2에 나타내었다. 평균과 표준편차를 이용하여 식 (1)과 같은 autoscaling에

Table 1. Archaeological data and symbols

Serial no.	Symbols	Items	Colors	Sites	Age
1		glassware	colorless	Hwangnam Great Tomb	5-6C(Silla)
2		"	"	Kyungju, Kyungbuk	"
3		"	blue	"	"
11		"	"	"	"
14		"	blue-black	"	"
4		glassbead	blue	Hwangnam Great Tomb	"
5		"	"	Kyungju, Kyungbuk	"
6		"	"	"	"
7		"	"	"	"
8		"	"	"	"
9		"	"	"	"
10		"	"	"	"
12		"	"	"	"
13	△	"	"	"	"
45		bead	blue	Imdangdong Tomb	5-6C(Kaya)
46		"	yellowish-green	Kyungsan, Kyungbuk	"
47		"	blue	"	"
48		"	"	"	"
49		"	yellowish-green	"	"
50		"	yellowish-green	Michu King's Tomb	5C(Silla)
51		"	"	Kyungju, Kyungbuk	"
52		"	blue	"	"
53		"	"	"	6C(Silla)

54	◇	◇	◇	◇
55	◇	◇	Bokhyundong Tomb, Taegu	5-6C(Silla)
56	◇	yellowish-green	◇	◇
57	◇	◇	◇	◇
58	◇	blue	Inwangdong Tomb Kyungju, Kyungbuk	5-6C(Silla)
59	knife ornament	blue	Hwangodong Tomb	5-6C(Silla)
60	glasswear	◇	Hwangnam Great Tomb	◇
79	bead	blue	Chonmachong	5-6C(Silla)
80	cup	colorless	◇	◇
84	bead	dark-blue	Hwangnam Great Tomb	◇
85	glasswear	blue	Kyungju, Kyungbuk	◇
89	bead	dark-blue	Joyangdong Tomb Kyungju, Kyungbuk	1-2C
90	◇	◇	Naedang Tomb, Taegu	5-6C(Kaya)
91	glass	light-green	◇	◇
92	bead	yellow	◇	◇
15	glassbead	blue	Okchun Tomb(No. 24)	4-5C
16	◇	◇	Hapchun, Kyungnam	◇
17	◇	◇	◇	◇
18	◇	◇	◇ (No. 9)	◇
19	◇	◇	◇	◇
20	◇	◇	◇	◇
21	◇	◇	◇	◇
22	◇	purple	◇ (No. 43)	◇
23	▲	◇	◇	◇
24	◇	◇	◇	◇
43	bead	blue	Jeopori Tomb	Kaya
44	◇	purple	Hapchon, Kyungnam	◇
61	◇	blue	Yangdong Tomb	2-3C
62	◇	◇	Kimhae, Kyungnam	◇
63	bead	yellowish-green	Pangyeje Tomb.	6C(Kaya)
64	◇	brown	Eichang, Kyungnam	◇
65	◇	◇	◇	◇
66	◇	◇	◇	◇
67	◇	◇	◇	◇
68	◇	yellowish-green	◇	◇
69	◇	◇	◇	◇
81	bead	greenish-blue	Tahori, Uichang, Kyungnam	BC 1C
82	◇	colorless	◇	◇
83	glass	colorless	Yangdong Tomb (1st) Kimhae, Kyungnam	2-3C
86	bead	orange	Yuldaeri Tomb (3rd)	5-6C(Kaya)
87	◇	dark-blue	Kosong, Kyungnam	◇
93	bead	blue	Pangyeje Tomb,	6C(Kaya)
94	◇	◇	Hapchon, Kyungnam	◇
41	glassbead	blue	Dogaedong Tomb	4-5C
42	◇	blue-black	Changwon, Kyungnam	◇
25	glass	green	Miruk temple, Iksan	6-7C(Paekche)
26	◇	◇	Chunbuk	◇

27		◇	◇	◇	◇
28		◇	◇	◇	◇
29		◇	◇	◇	◇
30		◇	◇	◇	◇
31		◇	◇	◇	◇
32	□	◇	◇	◇	◇
33		◇	◇	◇	◇
34		◇	◇	◇	◇
35		◇	◇	◇	◇
36		◇	◇	◇	◇
37		◇	◇	◇	◇
38		◇	◇	◇	◇
39		◇	◇	◇	◇
40		◇	◇	◇	◇
70	bead	orange	Muryong King's Tomb	6C(Paekche)	
71	◇	dark-red	Kongju, Chungnam	◇	
72	◇	dark-blue	◇	◇	
73	■	black	◇	◇	
74	◇	green	◇	◇	
75	◇	blue	◇	◇	
76	◇	orange	◇	◇	
77	glass	dark-red	◇	◇	
78	bead	black	◇	◇	
88	☆ bead	dark-blue	Oyari, Pyongyang	2-3(Laclang)	

Table 2. Mean values (ppm) and standard deviations

Element	Sm	Br	La	Ce	Lu	Th
Mean value (ppm)	1.89	15.3	11.2	42.7	0.233	5.06
Standard Deviation	1.61	13.6	10.3	60.7	0.167	5.03

Element	Cr	Hf	Sc	Fe	Co	Eu
Mean value (ppm)	33.3	5.15	3.18	$1.17 \times 10^4$	256	0.702
Standard Deviation	62.2	6.30	3.07	$1.55 \times 10^4$	369	0.823

의해 원인변수  $x_{ij}$ 를 신변수  $z_{ij}$ 로 변환시켰다. 식 (2)와 같은 주성분값  $y_{ij}$ 를 구하기 위해서는 신변수의 데이터 매트릭스  $[Z]$ 의 분산-공분산 매트릭스  $[R]$ 를 식 (3)에 의해 구하였다. 매트릭스  $[R]$ 로부터 식 (4), (5)의 과정을 통해 고유값  $\lambda_i$ 와 이에 대응

하는 고유벡터의 계수값을 순차적으로 구하였다. Table 3은 각 주성분의 고유값  $\lambda_i$ 와 전체 가변도에 대한 각 주성분의 기여도(%)를 나타내었으며, 주성분 1, 2, 3이 전체 가변도의 67%만을 차지함을 알 수 있다. 주성분 1, 2, 3에 대응하는 고유벡터의 계수값과 각 성분원소가 각 주성분에 기여하는 정도를 나타내는 공동성(communality)은 Table 4와 같다. 주성분 1에는 Sm, La, Th, Sc 및 Eu가 주성분 2에는 Co가 크게 기여함을 알 수 있다.

고유값  $\lambda_1, \lambda_2$ 에 대응하는 고유벡터의 계수값으로부터 식 (2)와 같이 주성분값  $y_{i1}, y_{i2}$ 를 구한 후 평면에 도시하면 Fig. 1과 같다. 시료는 입수된 지역에 따라 6개의 군을 이루고 있으며, 군 A, C(경북, 경남)가 서로 겹쳐 나타나고 있어 군의 분류가 불분명하다. Fig. 1은 전체 가변도의 54%만으로 도시된 것으로 46%의 가변도는 무시되고 있다.

이를 보완하기 위하여 주성분분석법에 의해 얻어진 주성분값의 전 가변도를 이용하는 비선형도시법으로 시료를 분류하였다. 이는 가변도가 큰 주성분값

Table 3. Eigenvalues and their contributions

Component	1	2	3	4	5	6	7
Eigenvalue	4.72	1.81	1.49	1.17	0.77	0.60	0.38
Contribution (%)	39.3	15.1	12.4	9.8	6.4	5.0	3.2
Cumulative (%)	39.3	54.4	66.8	76.6	83.0	88.0	91.2

Table 4. Eigenvector coefficients and communalities

Element	Eigenvector coefficient			Communality		
	Comp. 1	Comp. 2	Comp. 3	Comp. 1	Comp. 2	Comp. 3
Sm	0.353	0.169	-0.154	0.588	0.052	0.035
Br	0.214	0.121	0.445	0.216	0.026	0.295
La	0.336	-0.206	0.193	0.532	0.077	0.055
Ce	0.233	0.472	0.253	0.257	0.403	0.095
Lu	0.291	-0.130	-0.189	0.400	0.030	0.053
Th	0.369	-0.322	0.139	0.644	0.188	0.029
Cr	0.117	-0.122	0.247	0.065	0.027	0.091
Hf	0.292	-0.441	0.244	0.402	0.352	0.088
Sc	0.380	0.183	0.020	0.681	0.061	0.001
Fe	0.281	0.138	-0.463	0.372	0.035	0.319
Co	0.070	0.551	0.267	0.023	0.549	0.106
Eu	0.339	0.078	-0.464	0.544	0.011	0.320

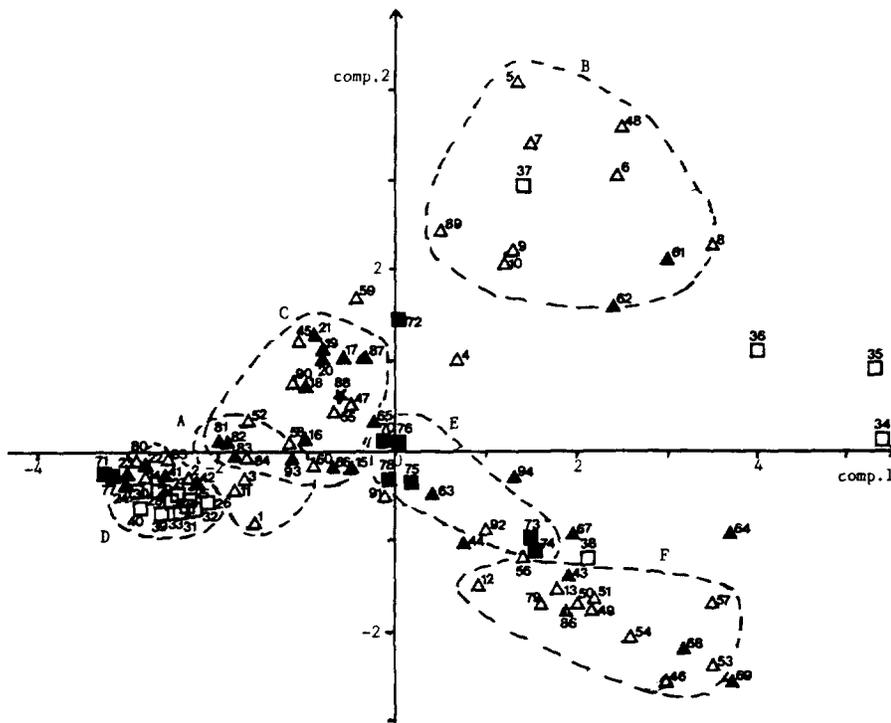


Fig. 1. The result of principal component analysis.

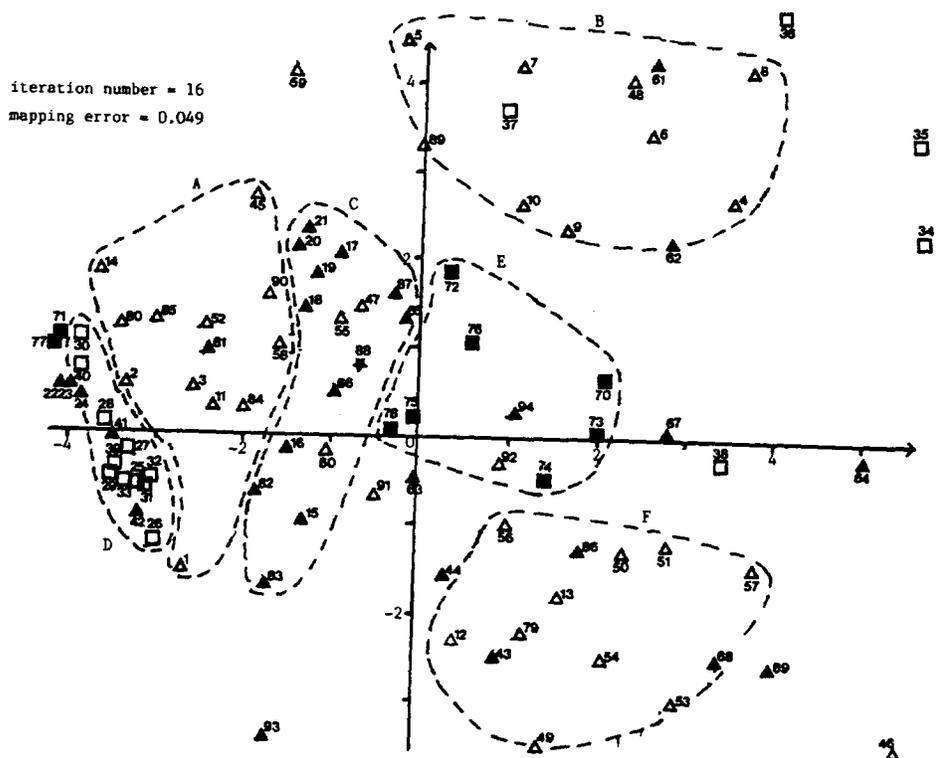


Fig. 2. Nonlinear mapping using two coordinates with largest variances.

1, 2를 기준으로 식 (6)에서 식 (10)의 과정을 반복하면서  $m$  차원의 공간에 나타내어진 주성분값을 2차원에 나타내도록 하였다. Iteration 수가 16이고 mapping 오차가 0.049일 때의 결과를 Fig. 2에 도시하였다. 시료의 분포가 넓게 나타나며 6개 군의 분류가 보다 분명함을 알 수 있다. 경북( $\Delta$ ) 지역 시료는 A, B, F의 3개군을 형성하였고, 경남( $\blacktriangle$ ), 전북( $\square$ ), 충남( $\blacksquare$ ) 지역 시료는 각각 별개의 군을 형성하였다.

시료를 입수한 위치와 주성분분석법 및 비선형도시법의 결과로 얻은 시료의 분포가 일치하면 SIMCA법을 위한 참조시료셋트(training set)로 삼고 그 외의 시료는 시험시료셋트(test set)로 삼아 시료의 분류를 재검토하였다.

선정된 6개의 참조시료셋트에 대하여 각각 식 (11)의 잔류오차  $e_{ik}^{(q)}$ 가 최소가 되도록  $\beta_{ik}^{(q)}$ ,  $\theta_{ik}^{(q)}$ 를 계산하였다. 이 때 항의 수  $A_q$ 를 구하기 위하여 항의 수를 하나씩 증가시키면서 구한 잔류오차의 제곱의 합인  $D_A$ 를 계산하였다. 이  $D_A$ 를 이용한 ( $D_{A-1} - D_A$ )

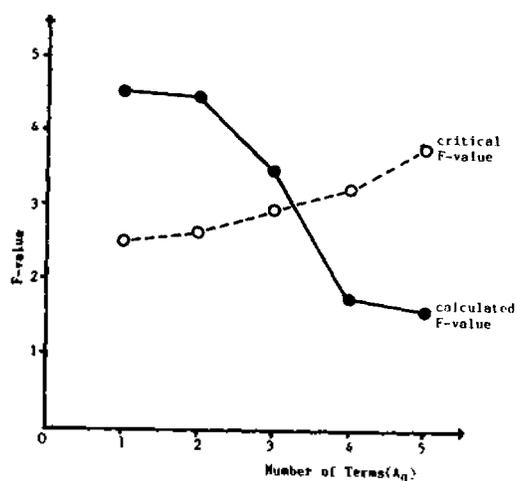


Fig. 3. Determination of the number of terms ( $A_q$ ) for the class E.

$(m - A_q)$  vs.  $D_A / (n_q - A_q - 1)(m - A_q)$ 의 값이  $(m - A_q)$  vs.  $(n_q - A_q - 1)(m - A_q)$ 에 대응하는 임계 F-값보다 작을 때의  $A_q$ 를 항의 수로 정하였다. 군 E (충남지역시료)의 경우를 Fig. 3에 나타내었다. 이

Table 5. Standard deviations and critical  $F$ -values of training set for each class. The numbers in parenthesis are the number of terms

Class ( $q$ )	A	B	C	D	E	F
Standard deviation, $S_o(q)$	0.3749	0.4690	0.3592	0.5294	0.5234	0.2425
Critical $F$ -value ( $A_p$ )	4.56(7)	3.43(5)	3.33(5)	3.36(5)	3.21(4)	4.74(7)

Table 6. Assignment of samples for training set and test set. The alphabet in parenthesis is the name of class

Sample no.	Symbols	Class given	Assignment	$F$ -value to nearest class	$F$ -value to related class
Training set for class A					
1		A	A	0.07 (A)	12.25 (C)
2		A	A	0.12 (A)	12.75 (C)
3		A	A	0.03 (A)	6.06 (B)
11		A	A	0.12 (A)	4.18 (C)
14		A	A	0.29 (A)	11.05 (C)
45		A	A	0.21 (A)	4.04 (B)
52	△	A	A	0.30 (A)	5.17 (C)
58		A	A	0.12 (A)	5.26 (C)
80		A	A	0.38 (A)	9.91 (C)
84		A	A, C	0.70 (A), 1.90 (C)	5.44 (B)
85		A	A	1.00 (A)	9.20 (C)
90		A	A, C	0.59 (A), 1.86 (C)	5.87 (B)
Training set for class B					
4		B	B	0.04 (B)	70.22 (A)
5		B	B	0.09 (B)	24.56 (E)
6		B	B	0.30 (B)	13.90 (E)
7	△	B	B	0.40 (B)	37.49 (C)
8		B	B	0.00 (B)	53.23 (C)
9		B	B	0.48 (B)	3.67 (E)
10		B	B	0.74 (B)	8.63 (E)
48		B	B	0.90 (B)	12.52 (E)
89		B	B, E	2.69 (B), 2.95 (E)	18.04 (C)
Training set for class C					
15		C	C	0.56 (C)	10.22 (A)
16		C	C	0.77 (C)	7.19 (B)
17		C	C	0.79 (C)	8.41 (A)
18		C	C	0.59 (C)	6.00 (B)
19		C	C	0.98 (C)	5.96 (B)
20		C	C	0.13 (C)	6.42 (B)
21	▲	C	C, A	0.33 (C), 3.64 (A)	7.18 (B)
65		C	C	0.50 (C)	4.57 (B)
66		C	C	0.81 (C)	5.07 (B)
82		C	C, A	0.18 (C), 2.79 (A)	5.94 (B)
83		C	C, A	0.34 (C), 2.72 (A)	6.58 (B)
87		C	C, E	0.02 (C), 2.74 (E)	4.88 (B)
Training set for class D					
25		D	D, A	1.13 (D), 1.12 (A)	7.18 (C)

26		D	D, A	0.53 (D), 2.27 (A)	6.33 (C)
27		D	D, A	0.25 (D), 0.80 (A)	8.41 (C)
28		D	D, A	0.13 (D), 0.47 (A)	9.29 (C)
29		D	D, A	0.26 (D), 0.86 (A)	8.09 (C)
30	□	D	D, A	0.45 (D), 0.21 (A)	12.28 (C)
31		D	D, A	0.29 (D), 1.20 (A)	7.16 (C)
32		D	D, A	0.11 (D), 1.71 (A)	7.19 (C)
33		D	D, A	0.37 (D), 0.99 (A)	7.77 (C)
39		D	D, A	0.49 (D), 0.87 (A)	8.32 (C)
40		D	D, A	0.94 (D), 0.41 (A)	11.44 (C)
Training set for class E					
70		E	E	0.35 (E)	5.03 (B)
72		E	E, B	0.00 (E), 2.02 (B)	9.28 (C)
73		E	E	0.09 (E)	13.43 (B)
74		E	E	0.16 (E)	18.51 (F)
75		E	E	0.18 (E)	5.42 (B)
76		E	E	0.43 (E)	8.65 (B)
78		E	E	0.75 (E)	4.61 (B)
Training set for class F					
12		F	F	0.36 (F)	5.55 (E)
13		F	F	0.00 (F)	12.92 (B)
49		F	F, B	0.25 (F), 0.05 (B)	21.63 (C)
50		F	F	0.05 (F)	9.00 (B)
51		F	F	0.08 (F)	10.07 (B)
53	△	F	F	0.11 (F)	14.23 (B)
54		F	F	0.75 (F)	11.64 (B)
56		F	F	0.31 (F)	5.32 (E)
57		F	F	0.04 (F)	19.86 (B)
79		F	F	0.01 (F)	13.93 (E)
Test set					
22		A, C, D	A	1.28 (A)	13.95 (C)
23		A, C, D	A	0.16 (A)	12.24 (C)
24		A, C, D	A, D	0.25 (A), 3.10 (D)	11.31 (C)
34		D	outlier		33.59 (B)
35		D	outlier		35.70 (B)
36		B, D	outlier		93.76 (B)
37		B, D	B	1.49 (B)	7.89 (E)
38		D	outlier		9.06 (B)
41		A, C, D	A, D	0.58 (A), 0.87 (D)	9.02 (C)
42		A, C, D	A	2.08 (A)	7.75 (C)
43		C, F	outlier		34.49 (F)
44		C, F	B	3.33 (B)	13.98 (C)
46		F	outlier		21.05 (B)
47		A, B, C	B	2.57 (B)	4.85 (A)
55		A, B, C	B	3.28 (B)	4.79 (A)
59		A, B	outlier		5.68 (A)
60		A, B, C	E	2.47 (E)	6.70 (C)
61		B, C	outlier		15.01 (B)
62		B, C	outlier		8.01 (B)

63	C, E, F	F	4.32 (F)	3.87 (B)
64	C	outlier		51.15 (F)
67	C, E	outlier		8.69 (E)
68	C, F	F	4.23 (F)	28.81 (B)
69	C, F	outlier		7.27 (F)
71	A, D, E	A	0.07 (A)	14.11 (C)
77	A, D, E	A	0.20 (A)	13.88 (C)
81	A, C	A, C	2.65 (A), 1.99 (C)	5.76 (B)
86	C, F	E	2.57 (E)	18.53 (F)
88	C	outlier		9.55 (C)
91	C, E, F	outlier		3.94 (E)
92	E, F	E	3.19 (E)	19.36 (F)
93	C, F	outlier		5.32 (B)
94	C, E	outlier		5.48 (E)

때는  $D_4$ 로부터 계산한  $F$ -값이  $A_4=4$ 일 때 임계  $F$ -값보다 적게 나타나므로 항의 수를 4로 정하였다. 이와 같은 방법으로 6개의 군에 대한 임계  $F$ -값과 항의 수를 구하였고, 식 (12)의 잔류오차를 이용한 각 군의 표준편차  $S_0^{(q)}$ 를 계산하여 Table 5에 나타내었다.

이와 같이 6개의 각 군에 대하여  $\beta_w^{(q)}$ ,  $\theta_{ab}^{(q)}$  및  $A_q$ 를 계산하여 각 군의 모델식을 만들고 이 모델식에 참조시료셋트와 시험시료셋트를 적용시켜 보았다. 이 결과를 Table 6에 나타내었으며 각 시료의 계산된  $F$ -값이 Table 5에 나타난 각 군의 임계  $F$ -값보다 적으면 시료는 관련군에 속하는 것으로, 임계  $F$ -값보다 크면 열외로 판정하였다. 주성분분석법과 비선형도시법의 결과로 선정된 각 군의 참조시료는 SIMCA법에 의해 관련군에 속하였으며, 열외 시료는 없었다. 군 A( $\Delta$ )에 속한 시료 84, 90은 군 C( $\blacktriangle$ )에도 속하였으며, 군 B( $\Delta$ )의 시료 89는 군 E( $\blacksquare$ )에도 속하는 것으로 나타났다. 군 C( $\blacktriangle$ )의 시료 21, 82, 83은 군 A( $\Delta$ )에, 시료 87은 군 E( $\blacksquare$ )에도 속하였다. 군 D( $\square$ )의 시료는 모두 군 A( $\Delta$ )와 동일한 군을 형성하였고, 군 E( $\blacksquare$ ), F( $\Delta$ )에 각각 속한 시료 72, 49는 군 B( $\Delta$ )에도 속하는 것으로 나타났다.

참조시료셋트에서 제외된 33개의 시험시료셋트를 각 군의 모델식에 적용시켜 본 결과 시료 22, 23, 42, 71, 77은 군 A( $\Delta$ )에 속하였으며, 시료 37, 44, 47, 55는 군 B( $\Delta$ )에 속하였다. 시료 60, 86, 92는 군 E( $\blacksquare$ )에 속하였고, 시료 63, 68은 군 F( $\Delta$ )에 속하는 것으로 나타났다. 또한 시료 24, 41은 군 A

( $\Delta$ )와 D( $\square$ )에 시료 81은 군 A( $\Delta$ )와 C( $\blacktriangle$ )에 동시에 속하는 것으로 나타났다. 이외의 16개 시료는 관련군으로 부터 모두 열외로 판정되었다.

본 연구에서는 unsupervised learning에 의해 6개 군을 형성함을 알 수 있었는데 이는 경북지역인 군 F( $\Delta$ )와 충남지역의 군 E( $\blacksquare$ )가 새로이 형성되었고 전보에서의 5개군 중 시료 34, 35, 36 및 38이 속한 군은 unsupervised learning의 결과 시료의 분포가 넓게 나타나므로 군으로 삼기가 어려워 제외했기 때문이다. 이들 시료는 SIMCA법의 시험시료로 삼아 속한 군을 판정한 결과 모두 열외로 판정되어 별개의 군으로 형성되어질 수 있음을 알 수 있었다. 또한 시료 12, 13은 전보에서 경북지역인 군 B( $\Delta$ )에 속하는 것으로 나타났으나 본 연구에서는 같은 경북지역인 군 F( $\Delta$ )와 더 유사하게 나타났다.

본 연구는 1990년 교육부 기초과학연구소 학술연구조성비(BSRI-90-339) 지원에 의해 이루어졌으며 이에 감사드립니다.

## 인 용 문 헌

1. D. L. Massart, B. G. M. Vandeginste, S. N. Deming, Y. Michotte, and L. Kaufman, "CHEMOMETRICS", Elsevier, Amsterdam, 1988.
2. S. Wold and M. Sjöström, "Chemometrics, Theory and Applications", B. R. Kowalski Ed., ACS symposium series, 52, 243 (1977).
3. K. Varmuza, "Pattern Recognition in Chemistry",

- Springer-Verlag, Berlin, 1980.
4. C. Lee, M. Z. Czae, S. Kim, H. T. Kang, and J. D. Lee, *Bull. Kor. Chem. Soc.*, **11**, 450 (1990).
  5. C. Lee, H. T. Kang, and S. Kim, *Bull. Kor. Chem. Soc.*, **9**, 223 (1988).
  6. C. Lee, O. C. Kwun, and H. T. Kang, *Bull. Kor. Chem. Soc.*, **7**, 73 (1986).
  7. C. Lee, S. Kim, and M. Y. Kim, *Bull. Kor. Chem. Soc.*, **9**, 5 (1988).
  8. J. Arunachalam and S. Gangadharan, *J. Indian Acad. Forensic Sci.*, **20**, 54 (1981).
  9. D. R. Preuss and P. C. Jurs, *Anal. Chem.*, **46**, 520 (1974).
  10. R. W. Liddell III and P. C. Jurs, *Anal. Chem.*, **46**, 2126 (1974).
  11. C. Lee, O. C. Kwun, I. C. Lee, and N. B. Kim, *J. Kor. Chem. Soc.*, **31**, 567 (1987).
  12. J. W. Sammon, Jr., *IEEE Trans. Comput.*, **C-18**, 401 (1969).
  13. S. Wold, *Pattern Recognition*, **8**, 127 (1976).
  14. J. I. Kim, *J. Radioanal. Chem.*, **63**, 121 (1981).
  15. C. Lee, *J. Kor. Nucl. Soc.*, **5**, 137 (1973).
  16. B. R. Kowalski and C. F. Bender, *J. Am. Chem. Soc.*, **95**, 686 (1973).