

## Scaling MDS for Preference Data Using Target Configuration<sup>1)</sup>

S.Y. Hwang<sup>2)</sup> · S.K. Park<sup>3)</sup>

### Abstract

MDS(multi-dimensional scaling) for preference data is a graphical tool which usually figures out how consumers recognize, evaluate certain products. This article is mainly concerned with an optimal scaling for MDS when target configuration is available. Rotation of axis and SUR(seemingly unrelated regression) methods are employed to get a new configuration which is obtained as close to the target as we can. Methodologies developed here are also illustrated via a real data set.

**Keywords** : MDS, Preference data, Rotation, SUR, Target configuration

### 1. 서 론

다차원 선호도 분석은 여러 상품들에 대한 개인 또는 그룹의 선호도를 알아보기 위한 분석 방법으로, MDS(multi-dimensional scaling)의 일종이라 할 수 있으며 결과는 보통 2차원 그림으로 제공된다. 즉, 선호 데이터는 주로 소비자(subjects)들의 제품(brands)에 대한 선호, 또는 제품 속성에 대한 각 제품별 평가를 말하는데, 이 기법은 이러한 두 종류의 대상을 동시에 평면상의 그림으로 표현해준다(임종원 et al.(2001)). 본 연구에서는 선호자료의 MDS 분석 시 제공되는 지각도(perceptual map) 및 MDS 그림 등을 그릴 때 생성된 좌표의 해석을 용이하게 하고, 결과의 변별력을 높이기 위해 기존의 자료로서의 목표좌표(target configuration)를 이용하여 (최적)좌표를 유도

- 
- 1) This work was supported by a research grant 2000 from Sookmyung Women's Univ.
  - 2) Professor, Department of Statistics, Sookmyung Women's Univ., Seoul, 140-742, Korea.  
E-mail : shwang@sookmyung.ac.kr
  - 3) Graduate student, Department of Statistics, Sookmyung Women's Univ., Seoul, Korea.  
E-mail : anova@sookmyung.ac.kr

하고자 한다.

여기서, 목표가 되는 좌표는 다음과 같은 세 가지 방법으로 얻을 수 있다. 예를 들면, 유사한 상품에 대한 외국의 조사결과, 국내의 다른 연구 결과, 시계열적 조사를 실시한 연구결과 등을 통해서 얻을 수 있다. 본 논문에서는 목표좌표에 자료를 근접 시킴으로써 해석의 편의를 도모하고자 하는 바, 외국의 조사결과 또는 국내의 다른 연구 결과를 이용하는 경우 유사한 상품의 심리적 위치를 비교해 볼 수 있으며, 시계열적으로 조사된 내용을 목표로 하는 경우 같은 상품이 시간에 따라 변화해 가는지의 추이도 알 수 있다.

따라서, 본 논문에서는 목표좌표를 이용한 최적좌표를 구하기 위해서 다음과 같이 구성을 하였다. 먼저 2장에서는 적용된 방법인 SUR(seemingly unrelated regression) 및 좌표의 회전을 간략히 살펴보고 3장에서는 2장에서 살펴본 방법들이 구체적으로 어떻게 적용되는지 알고리즘 제시를 통하여 알아보았으며, 4장에서는 앞서 논의한 바를 토대로 구체적인 사례를 통하여 살펴보았다.

## 2. 이용된 방법론의 소개

이미 살펴본바와 같이 선호자료의 MDS는 여러 상품들에 대한 개인(또는 그룹)의 선호도를 알아보기 위한 분석 방법으로 결과는 보통 2차원 그림으로 제공된다. 이 2차원 그림은 행렬도(biplot)의 일종으로서 Gabriel(1971)에 의해 정립된 것으로 자료행렬을 이차원 또는 저차원의 그래프로 표현하는 자료축약기법인 동시에 시각적인 기법이라 할 수 있다. 일반적으로 행렬도에서 사용되는 행렬은 본래 자료행렬 대신에 중심화·표준화된 행렬(X)로서, 비정칙값 분해(singular value decomposition : SVD)를 통해 나타나는 행렬들을 두 개의 행렬의 곱으로 나타내고 이를 이차원공간상에 근사시키게 된다. 행렬도를 구성하는 행(개체)의 그래프와 열(변수)의 그래프를 구하고 이를 하나의 그래프에 나타냄으로써 그들의 관계를 방향성을 통해 파악할 수 있으며 행과 열의 관계를 시각적으로 보여줌으로써 그들의 관계를 이해하는데 수월하다는 장점이 생기게 된다.

### 2.1 비정칙값 분해 및 행렬도

자료행렬  $X : n \times p$  의 계수가  $r$  일 때 열직교행렬(column orthogonal matrix)인  $U$  와  $V$  를 이용하여  $X$  를 다음과 같이 나타낼 수 있다.

$$X = UDV' = \sum_{k=1}^r u_k \lambda_k v_k' = \sum_{k=1}^r (u_k \lambda_k^c) (\lambda_k^{1-c} v_k') \quad (2.1)$$

여기서  $\lambda$  들은 비정칙값이다. 행좌표는  $UD^c$  으로 선택하고 열좌표는  $VD^{1-c}$  로 선택할 수 있으며  $c$  값에 따라 다양한 행렬도가 있을 수 있으나 주로 0, 0.5, 1인 경우를 사용한다. 좀더 자세한 내용은 (최용석(1999), 황선영 et al.(2001))을 참고하기 바란다.

## 2.2 SUR 모형

$m$  개의 종속변수  $y$  와  $r$  개의 독립변수들의 선형모형에 관해서 살펴보자. 예를 들면,

$$\begin{aligned} y_1 &= \beta_{01} + \beta_{11} x_1 + \cdots + \beta_{r1} x_r + \varepsilon_1 \\ y_2 &= \beta_{02} + \beta_{12} x_1 + \cdots + \beta_{r2} x_r + \varepsilon_2 \\ &\dots \\ y_m &= \beta_{0m} + \beta_{1m} x_1 + \cdots + \beta_{rm} x_r + \varepsilon_m \end{aligned}$$

와 같은 모형식을 회귀 방정식의 시스템이라고 한다. 모수  $\beta_i$ 의 추정은  $x_{it}$  와  $\varepsilon_{it}$ 의 상관관계 여부와  $\varepsilon = [\varepsilon_1, \varepsilon_2, \dots, \varepsilon_m]'$ 의 분산 공분산 행렬의 가정에 따라 크게 달라진다. SUR모형은 오차항이 동일시점(케이스)에만 상관관계를 가지고 있을 때의 회귀시스템이다. 즉, SUR 모형은 다음과 같이 표현된다.

$$\begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_m \end{pmatrix} = \begin{pmatrix} X_1 & O & \cdots & O \\ O & X_2 & \cdots & O \\ \vdots & \vdots & \ddots & \vdots \\ O & O & \cdots & X_m \end{pmatrix} \begin{pmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_m \end{pmatrix} + \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_m \end{pmatrix} \quad (2.2)$$

$\begin{matrix} Y & & & X & & \beta & & \varepsilon \end{matrix}$

이다. 여기서  $Y$ 는 크기가  $m(r+1)$ 인 열 벡터이며  $\text{Var}(\varepsilon) = \Omega$ 는  $\text{Cov}(\varepsilon_i, \varepsilon_j) = \sigma_{ij} I_T$ 이며  $\Omega$ 를 크로네커 곱(Kronecker product)  $\otimes$ 를 사용하여 표시하면  $\Omega = \Sigma \otimes I_T$ 가 된다. 물론  $\Sigma$ 는  $m \times m$  행렬로서  $(i, j)$  번째 원소는  $\sigma_{ij}$ 이다. SUR 모형에서의 추정량은

$$\begin{aligned} \hat{\beta}_{GLS} &= (X' \Omega^{-1} X)^{-1} X' \Omega^{-1} Y \\ &= (X' (\Sigma^{-1} \otimes I_T) X)^{-1} X' (\Sigma^{-1} \otimes I_T) Y \end{aligned} \quad (2.3)$$

이다. 만약,  $\Sigma$ 가 알려져 있다면 (2.3)의 추정량은 최소분산 불편 추정량임과 동시에 오차항이 정규분포를 따를 때의 ML추정량이 된다. 그러나 대부분의 경우  $\Sigma$ 가 알려져 있지 않으므로  $\Sigma$ 의 일치 추정량  $\hat{\Sigma}$ 를 이용한다.(cf. 박유성, 송석현(1998))

## 2.3. 회전행렬

MDS 분석은 자료로부터 개념상 의미 있고 해석하기 쉬운 속성과 개체들의 위치 등을 시각적으로 분석하는 것을 주목적으로 한다. 그러나 최초로 구해진 좌표는 때때로 해석하기 어려운 경우가 발생된다. 이러한 문제점을 해결하기 위하여 축(차원 ; 속

성)을 회전시키면 인자분석에서 인자의 단순구조를 얻을 수 있듯이 축이 단순한 구조로 변경이 가능하게 된다.  $\mathcal{C}$  을 실제 조사를 통해 구한 좌표 행렬이라고 하자.  $\mathcal{C}$  을 직교 회전시키기 위하여 직교변환 행렬  $T$ 를 이용한다. 이때  $TT' = T'T = I$ 를 만족하는 직교변환 행렬  $T$ 를 곱하여 얻어진 좌표 행렬은  $\mathcal{C}^* = \mathcal{C}T$ 가 된다. 회전각도  $\phi$ 와 회전방향에 따라 사용될 수 있는 직교변환 행렬  $T$ 는 다음과 같다.

$$T = \begin{bmatrix} \cos \phi & \sin \phi \\ -\sin \phi & \cos \phi \end{bmatrix} : \text{시계바늘 방향의 회전}$$

$$T = \begin{bmatrix} \cos \phi & -\sin \phi \\ \sin \phi & \cos \phi \end{bmatrix} : \text{시계바늘 반대방향의 회전}$$

### 3. 최적좌표의 선택

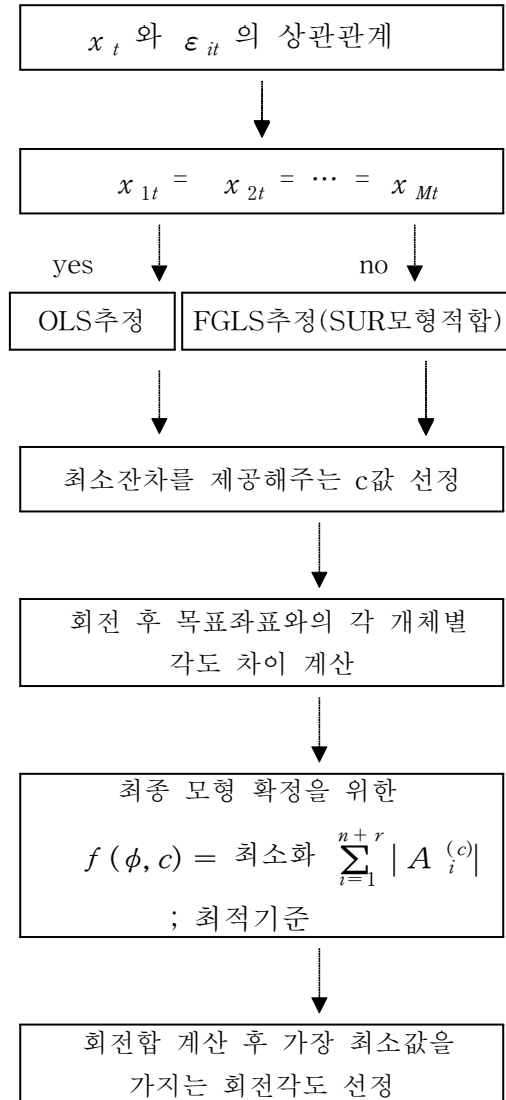
최적좌표 결정을 위한 첫 번째 단계는  $c$ 값을 결정하는 것이다. 최적의  $c$ 값을 선택하는 방법은 새로 생성된 좌표(  $0 \leq c \leq 1$  )의 두 차원을 열벡터(보통 2차원)로 보고 목표(target)좌표의 차원으로 회귀 시킨 후 모형에서 가장 최소의 잔차를 제공하는  $c$ 값을 선택하는 것이다. 여기서, 차원이(열)이 서로 독립일 경우 다변량 회귀(multivariate regression)를 사용하면 되지만, 대부분의 경우 차원끼리 상관관계가 있으므로 SUR 모형으로 적합하는 것이 적당하다. 구체적인 절차를 살펴보면 다음과 같다.

(1단계)  $0 \leq c \leq 1$  에서  $c$ 의 값을 순차적으로 변화시켜주면서 새로운 좌표를 생성한 뒤, 위의 좌표를 이용해서 만들어진 모형 중 최소 SSE값을 제공하는  $c$ 값을 선택한다.

(2단계) SUR모형 결과 선택되어진  $c$ 값을 가진 좌표를 일단 선정했다면 이제 목표(target)좌표와 가장 근접하도록 아래의 함수를 최소화 하는 회전각도  $\phi$ 를 선정한다.

$$f(\phi, c) = \text{최소화} \sum_{i=1}^{n+r} |A_i^{(c)}|$$

이 식에서  $A$  는 회전 후 각 좌표와 그에 대응하는 목표좌표 사이의 각도,  $\phi$  : 회전각도 그리고  $c$  는 (1단계)에서 구한 값을 나타낸다.



[ 그림 3-1 ] 최적좌표 선택 절차

#### 4. 사례분석 및 해석

본 절에서는 탄산음료를 즐겨 마시는 여대생 15명을 대상으로 음료의 선호도 및 대상들 간의 특성에 대한 설문 조사 자료에 앞서 서술한 방법을 적용시켜 보도록 하자. 6개 제품 (세븐업(7-up), 펩시(P), 코카콜라(CC), 닥터페퍼(DR), 슬라이스(S), 다이어트 콜라(DC))의 목표좌표 및 제품특성(T1,T2)은 성웅현(2000, p.369)에서 발췌하였으

며 6개 제품의 선호도 측정은 순위척도, 속성에 해당하는 맛의 강도(T1) 및 칼로리에 대한 느낌(T2)은 5점 척도(전혀 없음 1 - 매우 높음 5)로 평가하였다.

[ 표 4-1 ] 목표(target)좌표

	dim 1	dim2
7-up	-1.78	-0.89
P	1.20	0.05
CC	1.22	0.48
DR	0.61	-1.05
S	-0.26	-1.43
DC	0.78	0.85
T1	2.00	0.00
T2	0.00	2.00

목표좌표에서 좌표축의 의미를 살펴보면 첫 번째 좌표축은 맛의 강도(향의 정도)를 의미하고, 두 번째 좌표축은 칼로리의 정도(제품에 대한 주관적 느낌)를 의미한다고 판단된다. 이 결과에 의하면 Dr.Pepper와 Slice가 가장 높은 칼로리 청량음료 군이고, Diet Coke(코카콜라 라이트)가 가장 낮은 칼로리의 청량음료로 지각되고 있음을 알 수 있다. 또한 속성인 T1 및 T2를 차원으로 보았을 때, OLS 추정시 차원의 상관계수는 0.83033으로 상당히 높은 값을 보인다. 따라서 SUR 모형으로 적합하여,  $c$ 값 변화 ( $0 \leq c \leq 1$ )에 따른 잔차 제곱합의 변화를 통해 최적의  $c$ 값을 결정하도록 한다. [표 4-2]의 결과  $c = 0.1$ 인 경우에 조사된 자료로 생성된 좌표에 목표 좌표를 회귀시켰을 경우 가장 작은 SSE값을 제공하는 것으로 나타난다. 따라서  $c = 0.1$ 일 때의 좌표를 이용하여 회전을 실시 한 후, 목표 좌표와의 각 각도차이의 합을 구해 차이 각의 합이 최소화 될 때가 가장 목표와 근접한 그림을 제공하게 된다. [표 4.3] 결과로부터 생성된 좌표와 목표 좌표 사이의 각의 합이 가장 최소화 된 것은 회전 각도가 약 4.5 라디안 즉,  $260^\circ$  정도일 때이다.

회전 전 그림인 [그림 4.1]에서 속성벡터 의미를 살펴보면 첫 번째 벡터(T1)는 맛의 강도(향의 정도)를 의미하고, 두 번째 벡터(T2)는 칼로리의 정도(제품에 대한 주관적 느낌)를 의미한다고 판단된다. 이 결과에 의하면 크게 향의 정도는 Coke Classic과 Pepsi가 가장 높은 청량음료 군이고, Diet Coke가 가장 낮은 칼로리의 청량음료로 나타났다. 여기서, 일반적으로 생각할 때 대부분의 소비자들이 향이 강하고 칼로리가 높다고 생각되는 Coke Classic의 해석 및 7-Up, Slice, Dr. pepper의 음료 해석이 모호해진다. 따라서 MDS결과 속성이 직교하도록(해석이 편리)나온, 목표좌표와 가장 근접하게 회전한 [그림 4.2]를 보자. [그림 4.2]를 통해 회전 후의 속성벡터 의미를 살펴보면 첫 번째 벡터(T1)는 맛의 강도(향의 정도)를 의미하고, 두 번째 벡터(T2)는 칼로리의 정도(제품에 대한 주관적 느낌)를 의미한다고 판단된다. 이 결과에 의하면 크게 향의 정도는 Coke Classic과 Pepsi가 가장 높은 청량음료 군이고, 7-up과 Slice가 높은 칼로리의 청량음료 군으로 나타났고, Diet Coke와 Pepsi가 가장 낮은 칼로리의 청

량음료로 나타났다.

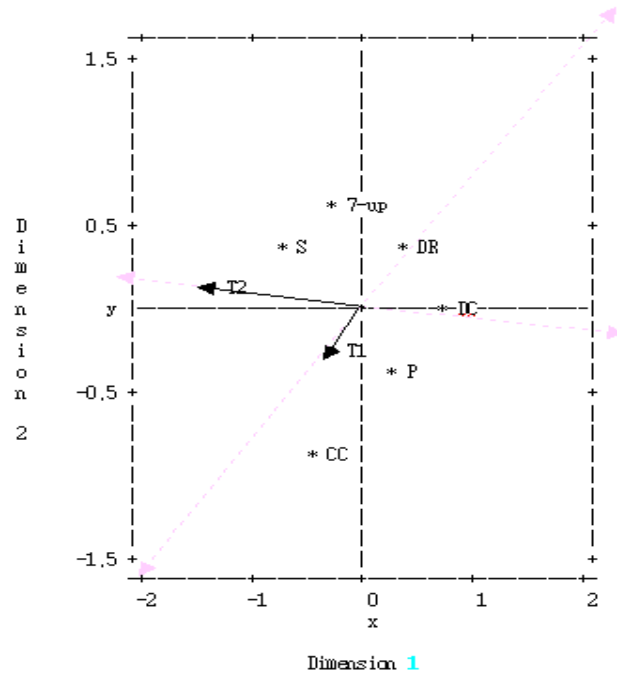
예제를 통한 논의 결과 회전 후의 그림이 속성(T2)에 있어서 해석에 편리함을 제공 해주었다. 하지만, 본 논문에서의 사례분석의 경우, 방법론을 제시하기 위하여 적은 표본 수와 상품을 이용했기 때문에 대표성 문제가 제기될 수 있을 것이다. 만약 추후 유사한 분석을 시도해 본다면, 결과 분석 전에 조사계획에서부터 실제 관측치를 얻기 까지 더 세심한 주의가 필요할 것이다. 또한 본 연구에서 조사된 속성 중 하나인 칼로리의 정도에 대한 개인의 주관적 느낌은 제품 자체에 칼로리의 표기가 명시된 경우가 많기 때문에 조사자들에게 혼란을 줄 여지가 있었다. 따라서, 이러한 속성도 더 깊이 생각해서 고려했다라면 좀더 좋은 결과를 제시해 줄 수 있을 것이라 예상된다.

[ 표 4 - 2 ] SUR 모형 적합시 SSE값의 변화

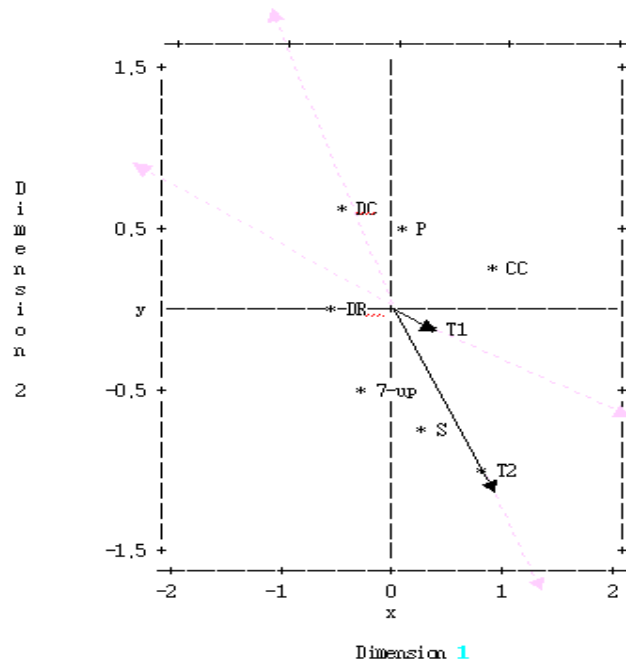
	SSE Y1	SSE Y2	TOTAL SSE = SSEY1 + SSEY2
C = 0.0	13.183668	15.942548	29.126216
<b>C = 0.1</b>	4.1465792	4.8126494	<b>8.9592287</b> ◀
C = 0.2	6.6899414	6.6899414	13.617698
C = 0.3	10.374306	11.570306	21.944612
C = 0.4	14.91594	17.384623	32.300563
C = 0.5	25.76496	26.016992	51.781952
C = 0.6	40.758051	42.994519	83.75257
C = 0.7	112.3969	66.808166	179.20506
C = 0.8	101.02285	103.8407	204.86354
C = 0.9	159.97221	160.49563	320.46784
C = 1.0	251.99159	247.64513	499.63672

[표 4 - 3] 회전 후 각도 차이의 합

회전각도( $\phi$ )	0.5 ( 28.6°)	1.0 ( 57.3°)	1.5 ( 85.9°)
$\sum_{i=1}^{n+r}  A_i^{(c)} $	10.4424	13.0786	13.78777
회전 각도( $\phi$ )	2.0 ( 114.6°)	2.5 ( 143.2°)	3.0 ( 171.9°)
$\sum_{i=1}^{n+r}  A_i^{(c)} $	12.9855	11.8394	9.8398
회전 각도( $\phi$ )	3.5 ( 200.5°)	4.0 ( 230°)	<b>4.5 ( 258°)</b>
$\sum_{i=1}^{n+r}  A_i^{(c)} $	7.8392	6.0547	<b>5.0637</b>
회전 각도( $\phi$ )	5.0 ( 286.5°)	5.5 ( 315.1°)	6.0 ( 343.8°)
$\sum_{i=1}^{n+r}  A_i^{(c)} $	5.7692	6.4438	8.4438



[ 그림 4-1 ] 회전 전 plot



[ 그림 4-2 ] 회전 후 plot



## 참 고 문 헌

1. 박유성, 송석현 (1998). *SAS/ETS를 이용한 경영경제자료 분석*. 정일출판사.
2. 성웅현 (2000). *응용다변량 분석 : 이론과 SAS의 활용*. 탐진출판사.
3. 임종원, 박형진, 강명수 (2001). *마케팅 조사 방법론*. 법문사.
4. 최용석 (1999). *행렬도의 이해와 응용*. 부산대학교출판부.
5. 황선영, 정수진, 김영원(2001). 다차원선호분석의 최적척도화 및 부분수량화. *응용통계 연구*, 14권, 305-320.
6. Gabriel, K.R.(1971). The biplot graphic display of matrices with applications to principal component analysis. *Biometrika*, 58, 453-467.

[ 2003년 3월 접수, 2003년 4월 채택 ]