

독립성분분석을 이용한 혼합물의 미지성분비율 예측

이혜선¹⁾ 송재기²⁾ 박해상³⁾ 전치혁⁴⁾

요약

독립성분분석은 차원이 높은 다변량데이터로부터 기저구조를 형성하는 독립성분을 분리하는데 사용되는 기법으로서 패턴인식, 예측 등 2차적 분석을 위한 1차 분석단계에서 사용할 수 있다. 본 연구에서는 독립성분분석을 이용하여 여러 혼합물 데이터로부터 독립성분을 분리한 다음 각 구성성분의 혼합비율을 예측하는 절차를 제안한다. 적용으로서 도금강판의 엑스선 회절강도값으로부터 여러가지 상을 분리한 다음 비음최소자승법을 이용하여 각 상의 분율을 예측하였으며, 이러한 제안방안의 타당성 평가를 위하여 모의실험을 실시하였다.

주요용어: 독립성분분석, 비음최소자승법, 비정규성

1. 서론

독립성분분석(Independent component analysis; ICA)은 다변량 데이터로부터 서로 독립인 성분을 분리하는데 사용되는 통계적 기법이다. ICA는 주성분분석(Principal component analysis), 요인분석(Factor analysis), Projection pursuit 등과 같이 일종의 선형변환방법이라고 볼 수 있으며, 최근 영상의학, 통신, 이미지분석 등 다양한 분야에서 기저구조를 밝혀내기 위한 기법으로 적용되고 있다 (Ikeda and Toyama, 2000; Hyvärinen et al., 1998). 유사한 목적으로 NMF (Nonnegative matrix factorization) 등이 사용되기도 한다 (Lee and Seung, 1999, 2001).

본 연구의 목적은 ICA를 이용하여 여러가지 순수물질이 섞인 혼합물에 대하여 측정된 시그널 데이터로부터 구성성분들을 도출하고, 그 독립성분들의 구성비율을 비음최소제곱법(Non-negative least square)에 의해서 예측하는 것이다. 본 연구에서 제안하는 방법은 도금강판 표면의 엑스선회절(X-ray Diffraction) 강도값을 이용한 각 상의 비율예측에 적용되었으며, 제안방법의 타당성을 평가하기 위해 모의실험을 실시하였다.

1) (702-701) 대구광역시 북구 산격동 1370번지 경북대학교 통계학과, 박사과정

E-mail : hyelee@postech.ac.kr

2) (702-701) 대구광역시 북구 산격동 1370번지 경북대학교 통계학과, 교수

E-mail:jksong@knu.ac.kr

3) (790-784) 경북 포항시 남구 효자동 산 31 포항공과대학교 산업경영공학과, 석박사통합과정

E-mail : shoo359@postech.ac.kr

4) (790-784) 경북 포항시 남구 효자동 산 31 포항공과대학교 산업경영공학과, 교수

E-mail : chjun@postech.ac.kr

2. 독립성분분석의 소개

2.1. 개요

ICA 방법은 데이터로부터 서로 독립인 성분을 분리해내는 미지성분분리(blind source separation) 기법이라고도 한다. 즉, p 개의 서로 독립인 성분이 다양한 비율로 혼합된 m 개의 대상 (이를 혼합물(mixture)이라 함)에서 관측된 데이터로부터 p 개의 독립인 성분을 분리하고자 하는 것이다. 단, 여기서 $m \geq p$ 이어야 한다. 예를 들어서 어떤 파티장에 p 개의 음원(예를 들어 배경음악, 목소리 등)이 있고 그 안의 여러위치에 설치된 m 개의 마이크로 녹음을 하여 파장데이터 (n 개의 주파수에 대한 진폭)를 얻었다고 할 때, 이 $(m \times n)$ 차원의 혼합물 데이터를 바탕으로 p 개의 독립된 음원을 분리하는 것이다 (Hyvärinen, 1999; Hyvärinen and Oja, 2000). S_i 를 i 번째 ($i = 1, \dots, p$) 성분을 나타내는 확률변수라 할 때 j 번째 혼합물 변수인 X_j 는 다음과 같이 p 개의 성분결합으로 이루어진다.

$$X_j = a_{j1}S_1 + a_{j2}S_2 + \dots + a_{jp}S_p = \sum_{i=1}^p a_{ji}S_i \quad (j = 1, \dots, m) \quad (2.1)$$

여기서 a_{ji} 는 i 번째 독립성분에 대한 j 번째 혼합물계수를 나타낸다. X_{jk} 를 j 번째 혼합물에 대한 k 번째 관측치, S_{jk} 를 j 번째 독립성분에 대한 k 번째 관측치라 하면 ($k = 1, \dots, n$), $\mathbf{X} = (X_{jk})$, $\mathbf{S} = (S_{jk})$ 일 때 식 (2.1)를 다음과 같이 표현할 수 있다.

$$\mathbf{X} = \mathbf{A}\mathbf{S} \quad (2.2)$$

여기서 $\mathbf{A} = (a_{ji})$ 는 $(m \times p)$ 행렬로서 혼합물계수행렬이라한다. ICA는 관측된 \mathbf{X} 로부터 S_i 들이 서로 통계적으로 독립이 되어야 한다는 제약조건에서 \mathbf{A} 와 \mathbf{S} 로 분리하는 것이라 할 수 있다.

2.2. 비정규성척도

식 (2.1)에서 하나의 성분을 나타내는 확률변수 S (편의상 아래첨자 생략)는 다음과 같은 형태로 나타낼 수 있다.

$$S = w_1X_1 + \dots + w_mX_m = \mathbf{w}^T \mathbf{x} \quad (2.3)$$

여기서 $\mathbf{w}^T = (w_1, \dots, w_m)$ 은 가중치 벡터이며 $\mathbf{x}^T = (X_1, \dots, X_m)$ 은 혼합물 변수벡터이다. 따라서 ICA에서는 독립적인 S 를 도출하기 위하여 $\mathbf{w}^T \mathbf{x}$ 의 비정규성을 최대화시키는 가중치 벡터 \mathbf{w} 를 찾고자 한다(Hyvärinen and Oja, 2000). 여기서 비정규성 척도로 negentropy를 사용하는데 다음과 같이 정의된다.

확률변수 벡터 \mathbf{y} 의 확률밀도함수가 $f(\mathbf{y})$ 일 때 이에 대한 entropy는 다음과 같이 정의된다.

$$H(\mathbf{y}) = - \int f(\mathbf{y}) \log f(\mathbf{y}) d\mathbf{y} \quad (2.4)$$

\mathbf{y} 가 정규분포를 따를 때 동일한 분산을 갖는 확률변수 중 위의 entropy가 최대가 됨을 이용하여 negentropy J 를 다음과 같이 정의한다.

$$J(\mathbf{y}) = H(\mathbf{y}_{gauss}) - H(\mathbf{y}) \quad (2.5)$$

위에서 \mathbf{y}_{gauss} 는 \mathbf{y} 와 동일한 분산-공분산 행렬을 갖는 정규 확률변수벡터이다. 즉, 위의 negentropy는 \mathbf{y} 가 정규분포를 따를 때 0이 되며 다른 분포를 따를 때는 양의 값을 가진다. 비정규성척도인 negentropy J 를 최대화함으로써 비정규성을 띠게 된다.

ICA에서는 negentropy를 근사시키기 위하여 다음과 같은 식을 사용한다.

$$J(\mathbf{y}) \approx \{E[G(\mathbf{y})] - E[G(\mathbf{z})]\}^2 \quad (2.6)$$

여기서 \mathbf{z} 는 평균 0, 분산-공분산행렬이 단위행렬(\mathbf{I})인 정규확률변수를 의미하며, 함수 G 로 는 임의의 non-quadratic 함수를 사용할 수 있으나 보통 아래와 같은 함수를 사용한다.

$$G_1(u) = \frac{1}{a_1} \log \cosh(a_1 u) \quad (2.7a)$$

$$G_2(u) = -\exp(-u^2/2) \quad (2.7b)$$

식 (2.7 a)에서 a_1 은 주로 $1 \leq a_1 \leq 2$ 에서 취해지며, $a_1 = 1$ 이 자주 사용된다 (Hyvärinen and Oja, 2000).

2.3. FastICA 알고리즘

널리 사용되는 FastICA는 식 (2.3)에서의 가중치벡터 \mathbf{w} 를 구하기 위하여 식 (2.6)의 negentropy를 최대화시킨다. 즉,

$$\max_{\mathbf{w}} J(\mathbf{w}^T \mathbf{x}) \quad (2.8a)$$

$$\text{subject to } \|\mathbf{w}\| = 1 \quad (2.8b)$$

위의 최적화문제 해결을 위해 고정점알고리즘을 이용한 Newton 방법을 사용한다. 결국 하나의 \mathbf{w} 를 구하기 위하여는 다음과 같은 과정이 진행된다.

Step 1. \mathbf{w} 의 초기값 선택

Step 2. $\mathbf{w} \leftarrow E[\mathbf{x}g(\mathbf{w}^T \mathbf{x})] - \mathbf{w}E[g'(\mathbf{w}^T \mathbf{x})]$ (기대치는 데이터의 평균으로 추정)

여기서 함수 g 는 non-quadratic 함수 G 의 미분함수이다.

Step 3. $\mathbf{w} \leftarrow \mathbf{w}/\|\mathbf{w}\|$

Step 4. 수렴하면 stop. 그렇지 않으면 Step 2 반복

알고리즘의 유도 및 상세과정은 Hyvärinen and Oja(2000)를 참조할 수 있다.

2.4. 미지성분 비율예측

위의 알고리즘으로부터 w 를 구한 후 식 (2.3)을 이용하여 각 독립성분을 얻을 수 있으며 다시 식 (2.1)의 모형을 통하여 혼합물 계수 a_{ji} 를 산출할 수 있다. 그러나 본 연구에서와 같이 a_{ji} 가 j 번째 혼합물의 i 번째 성분의 혼합비율을 나타내는 경우에는 비율의 계수가 되어야 한다. 따라서 $a_{ji} \geq 0$ 이고 $\sum a_{ji} = 1$ 인 제한식이 적용된 비음최소제곱법(Non-negative least squares method)을 적용한다. 즉, j 번째 시료에 대하여 사용되는 회귀모형은 아래와 같다.

$$X_{jk} = \sum_{i=1}^p a_{ji} S_{ik} + \varepsilon_{jk} \quad (k = 1, \dots, n) \quad (2.9)$$

단, $a_{ji} \geq 0$, $\sum a_{ji} = 1$ 이며, ε_{jk} 는 회귀모형관련 오차항을 나타낸다. 비음최소제곱법은 Lawson과 Hanson(1974, p.161)의 알고리즘을 구현한 Matlab의 LSQNONNEG(Linear least squares with nonnegativity constraints) 함수를 이용한다. 이상의 과정을 도식화하면 그림 2.1과 같다.

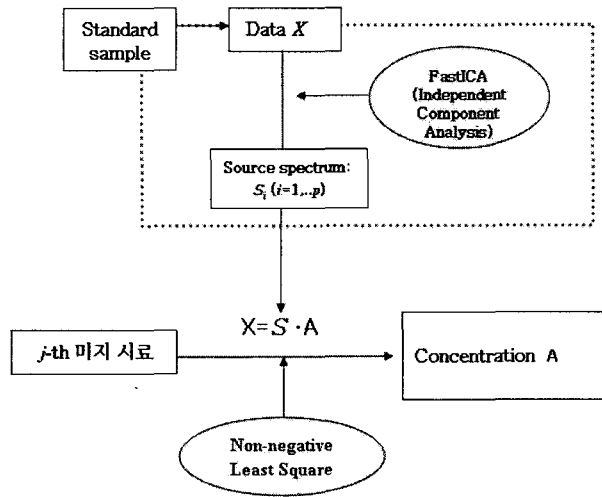


그림 2.1: ICA와 비음최소제곱법에 의한 성분비율 예측과정

3. 도금강판 표면 상(phase)별 분을 예측

본 연구에서는 도금강판의 엑스선 회절강도값을 이용하여 구성하고 있는 여러가지 상(phase)에 대한 구성비율을 예측하고자 한다. 엑스선 회절법은 물질의 구조를 밝혀내는데 적용되는 분광기술이다. 금속의 회절강도 패턴으로부터 미지시료내의 물질을 확인할 수 있고 혹은 이미 확인된 물질의 분자구조를 분석할 수 있다. 본 연구에서 사용된 시료는 네가지 상(phase)- 에타, 제타, 델타, 감마상-으로 구성되어 있고, 각 상의 피이크 (peak)는

어느 각도에서 나오는지 알려진 것이다. 상의 분율을 구하는 방법으로 기존에는 회절강도 값의 피이크의 규모에 따라 적분방식을 사용하는데 측정조건에 따라 피이크의 스케일이 달라지므로 일관성있는 결과를 얻을수 없다. 한편, 그림 3.1의 (b)와 같이 표면의 이미지 사진을 찍어서 이미지 분석을 하는 방안도 고려할 수 있으나 각 상의 모양의 구분과 영역이 뚜렷하지 않아 어렵다는 것이다. 이에 따라 ICA를 이용하여 성분인 네가지 상의 벡터를 얻고 원래의 엑스선 회절강도값에 회귀시키는 방식으로 각 분율을 예측하게 되었다.

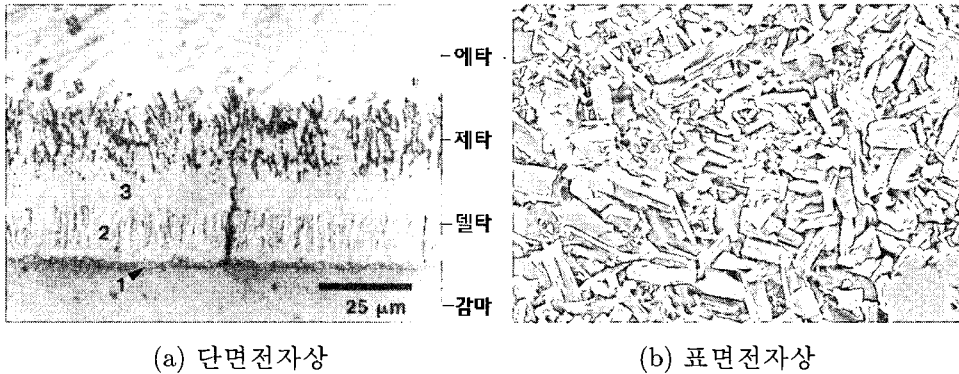


그림 3.1: 금속의 단면과 표면상

본 연구에서 사용한 20개 시료에 대한 스펙트럼 그림 3.2은 68도에서 78.8도까지 0.05도 간격의 216관측차원(x축)에서 측정된 엑스선회절강도값(y축)이다.

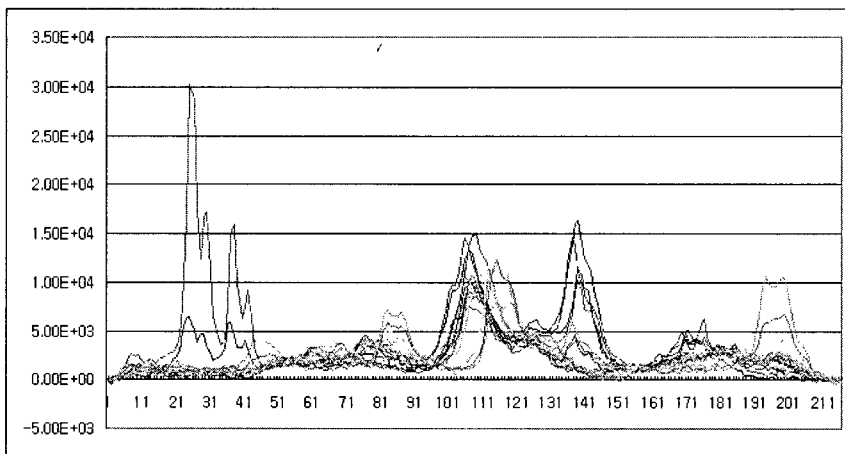


그림 3.2: 엑스선 회절강도 스펙트럼

이로부터 ICA를 이용하여 얻은 독립성분 스펙트럼은 그림 3.3에서 보는 바와 같이 5가

지로 분리되었으며 이 중 델타상이 두 개의 상으로 구성되어 있음이 확인되었다. 다른 여러 개의 실험데이터로부터 일관성있게 그림 3.3의 독립성분이 얻어짐으로써 ICA 적용의 가능성이 확인되었다. 분석프로그램은 R(version1.9)과 Matlab을 사용하였다.

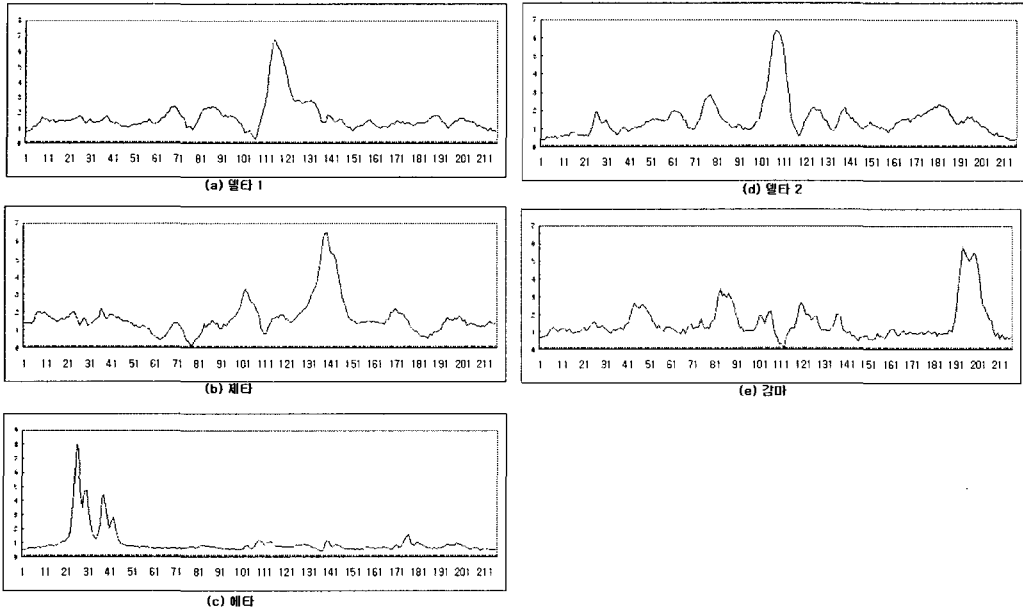


그림 3.3: 추출된 5가지 상별 독립성분

다음으로 2절에서 제안한 방안을 적용하여 상별 분율을 예측하였으며 이의 결과를 표 3.1과 같이 정리하였다. 여기서 델타상은 그림 3.3에서 보듯이 두가지로 나타났으나 분율산출에서는 이를 합하였다.

도금강판의 상별 분율은 실제로 측정할 수 없는 값이므로 이의 적합성을 검증하기는 어렵다. 본 결과는 생산현장 전문가에 의뢰하여 그들이 갖고있는 간접적인 정보를 활용하여 타당성을 확인하였다.

4. 모의실험

본 제안방법의 타당성을 검증하기 위한 모의실험을 수행한다. 실제독립성분을 아는 것으로 가정하고 성분비율을 다양하게 혼합한 관측치를 생성한 후 제안한 방법에 의거한 성분비를 추정치를 실제비율과 비교하고자 한다.

4.1. 독립성분의 생성

모의실험을 위해 그림 4.1과 같은 세가지 종류의 독립성분을 가정하였는데 여기서 S_1 은

표 3.1: 도금강판의 예측된 성분비율

시료번호	감마	제타	에타	델타
1	0.00	0.46	0.10	0.44
2	0.00	0.11	0.00	0.89
3	0.00	0.00	0.00	1.00
4	0.00	0.00	0.00	1.00
5	0.00	0.25	0.64	0.11
6	0.20	0.41	0.00	0.59
7	0.40	0.80	0.00	0.90
8	0.60	0.00	0.00	0.96
9	0.00	0.00	0.00	0.94
10	0.00	0.16	0.00	0.84
11	0.00	0.34	0.00	0.66
12	0.00	0.18	0.56	0.26
13	0.00	0.27	0.00	0.73
14	0.10	0.00	0.00	0.99
15	0.60	0.00	0.00	0.94
16	0.10	0.00	0.00	0.90
17	0.18	0.00	0.00	0.82
18	0.38	0.00	0.00	0.62
19	0.61	0.00	0.00	0.39
20	0.82	0.00	0.00	0.18

피크가 가장 뒤쪽에 있도록 하고, S_2, S_3 은 일부 피크가 서로 겹치는 부분이 존재하도록 하였다. 여기서 x축은 관측차원, y축은 각 독립성분의 수준을 나타낸다.

4.2. 독립성분 혼합비율

다양한 혼합물을 생성하기 위하여 세가지 독립성분 S_1, S_2, S_3 에 대한 실제 혼합비율은 표 4.1과 같으며 다음과 같이 생성하였다. 총 20개의 시료로 구성되는데 9번째 관측치까지는 S_1 은 없이 S_2, S_3 구성비율을 0.1씩 증가(감소)시켰고, 10번째 관측치부터는 S_1 에 대해 0.9부터 0.1씩 감소시키고 남은 비율은 S_2, S_3 의 시료 1~9번째까지의 비율대로 할당하였다. 19번째와 20번째에는 S_3 없는 두가지 혼합비율을 포함시켰다.

혼합물데이터 \mathbf{X} 는 다음의 식으로부터 생성하였다. 여기서 오차항 ε_{jk} 은 평균이 0이고, 세가지 수준(잡음수준)의 표준편차 (0.01, 0.05, 0.1)를 가진 정규분포로부터 생성한다.

$$X_{jk} = \sum_{i=1}^3 a_{ji} \times S_{ik} + \varepsilon_{jk}, \quad j = 1, \dots, 20; k = 1, \dots, 20 \quad (4.1)$$

한 예로 그림 4.2는 시료 14번에 대한 것으로 S_1, S_2, S_3 가 각각 0.5, 0.25, 0.25로 혼합된 것이며, 오차항 표준편차(잡음수준)를 각각 0.01, 0.05, 0.1로 하였을 때 생성된 혼합물 스펙트럼을 보여준다. x축은 관측차원, y축은 독립성분의 수준이다.

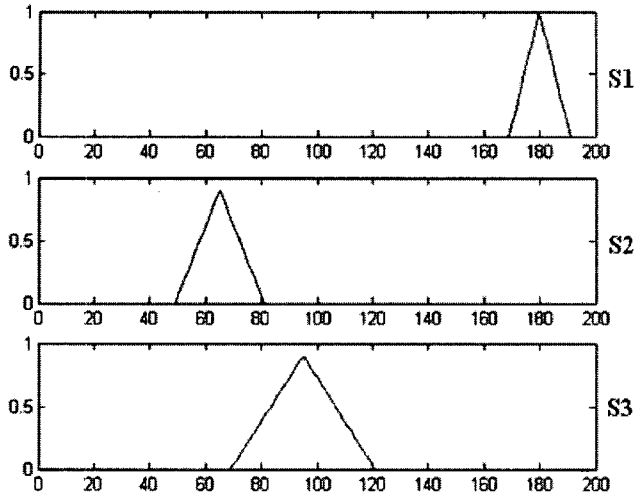


그림 4.1: 가정된 독립성분

표 4.1: 시료별 실제 성분비율

표본	S_1 비율	S_2 비율	S_3 비율
1	0.00	0.10	0.90
2	0.00	0.20	0.80
3	0.00	0.30	0.70
4	0.00	0.40	0.60
5	0.00	0.50	0.50
6	0.00	0.60	0.40
7	0.00	0.70	0.30
8	0.00	0.80	0.20
9	0.00	0.90	0.10
10	0.90	0.01	0.09
11	0.80	0.04	0.16
12	0.70	0.09	0.21
13	0.60	0.16	0.24
14	0.50	0.25	0.25
15	0.40	0.36	0.24
16	0.30	0.49	0.21
17	0.20	0.64	0.16
18	0.10	0.81	0.09
19	0.20	0.80	0.00
20	0.40	0.60	0.00

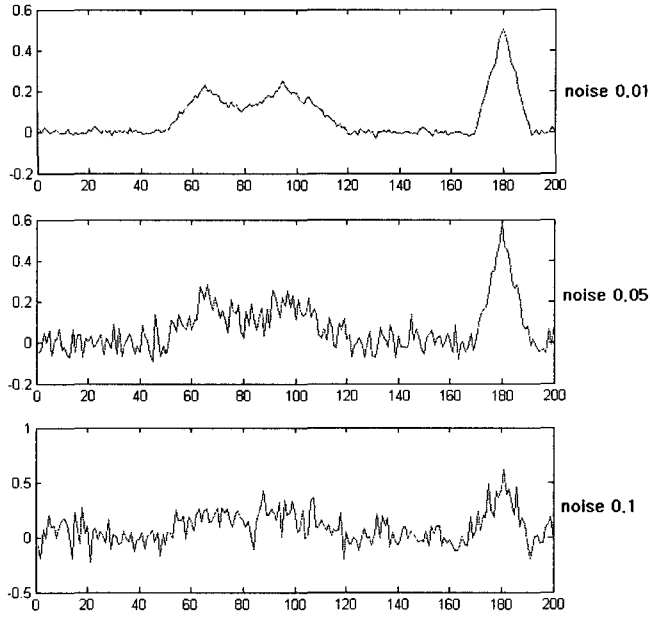


그림 4.2: 잡음 수준에 따른 혼합물 스펙트럼 (14번 시료)

4.3. 모의실험 결과

ICA 알고리즘을 적용하여 얻은 독립성분은 그림 4.3과 같으며, 잡음수준에 따라 다소 차이가 있으나 원래의 가정된 독립성분인 그림 4.1과 유사한 형태를 가짐을 볼 수 있다. 여기서 x축은 관측차원이며, y축은 독립성분의 수준이다.

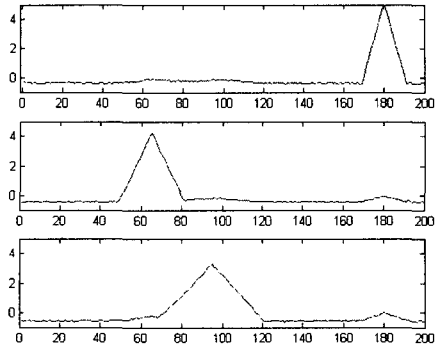
그림 4.3의 세가지 독립성분을 비음최소제곱법에 적합시켜 추정된 구성비율은 표 4.2와 같다. 표 4.2의 (b)-(d)의 마지막 열은 각 독립성분의 실제비율(a_{ji})과 예측비율(\hat{a}_{ji})의 차이의 절대값의 합을 나타낸다.

$$\text{절대오차의 합} = \sum_{i=1}^3 |a_{ji} - \hat{a}_{ji}| \tag{4.2}$$

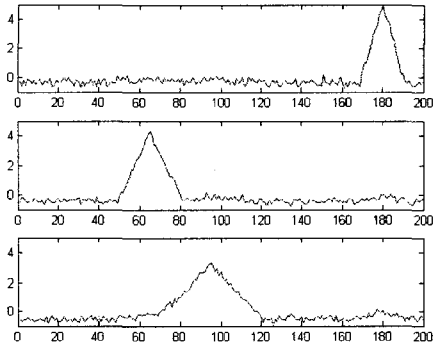
본 논문에서는 성분비율의 전반적인 예측정확도를 비교하는 척도로서 평균절대오차와 평균제곱오차를 사용하였다.

$$\text{평균절대오차} = \frac{1}{60} \sum_{j=1}^{20} \sum_{i=1}^3 |a_{ji} - \hat{a}_{ji}| \tag{4.3}$$

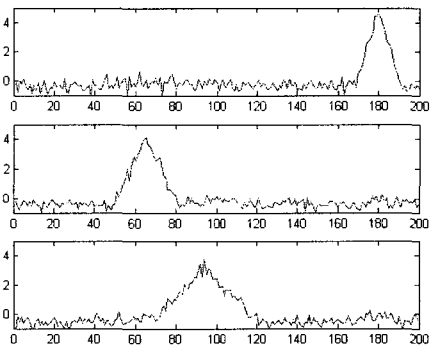
$$\text{평균제곱오차} = \sqrt{\frac{1}{60} \sum_{j=1}^{20} \sum_{i=1}^3 (a_{ji} - \hat{a}_{ji})^2} \tag{4.4}$$



(a) 잡음수준 0.01



(b) 잡음수준 0.05



(c) 잡음수준 0.1

그림 4.3: ICA에 의한 독립성분 (잡음수준별)

표 4.2: 예측 혼합비율의 절대값오차

(a) 실제 혼합비율

표본	S_1 비율	S_2 비율	S_3 비율
1	0.00	0.10	0.90
2	0.00	0.20	0.80
3	0.00	0.30	0.70
4	0.00	0.40	0.60
5	0.00	0.50	0.50
6	0.00	0.60	0.40
7	0.00	0.70	0.30
8	0.00	0.80	0.20
9	0.00	0.90	0.10
10	0.90	0.01	0.09
11	0.80	0.04	0.16
12	0.70	0.09	0.21
13	0.60	0.16	0.24
14	0.50	0.25	0.25
15	0.40	0.36	0.24
16	0.30	0.49	0.21
17	0.20	0.64	0.16
18	0.10	0.81	0.09
19	0.20	0.80	0.00
20	0.40	0.60	0.00

(b) 예측 혼합비율(잡음 수준:0.01)

표본	S_1 예측	S_2 예측	S_3 예측	\sum	오차
1	0.00	0.02	0.98	0.05	
2	0.00	0.12	0.88	0.05	
3	0.00	0.23	0.77	0.05	
4	0.00	0.34	0.66	0.04	
5	0.00	0.45	0.55	0.04	
6	0.00	0.56	0.44	0.03	
7	0.00	0.68	0.32	0.01	
8	0.00	0.81	0.19	0.00	
9	0.00	0.94	0.06	0.02	
10	0.93	0.00	0.07	0.02	
11	0.83	0.00	0.17	0.03	
12	0.72	0.04	0.24	0.03	
13	0.60	0.12	0.28	0.03	
14	0.48	0.23	0.29	0.03	
15	0.37	0.35	0.28	0.02	
16	0.26	0.52	0.23	0.03	
17	0.15	0.69	0.16	0.04	
18	0.04	0.90	0.06	0.06	
19	0.15	0.85	0.00	0.03	
20	0.37	0.63	0.00	0.02	

(c) 예측 혼합비율 (잡음 수준:0.05)

표본	S_1 비율	S_2 비율	S_3 비율	\sum	오차
1	0.00	0.03	0.97	0.05	
2	0.00	0.13	0.87	0.05	
3	0.00	0.21	0.79	0.06	
4	0.00	0.34	0.66	0.04	
5	0.00	0.43	0.57	0.04	
6	0.00	0.53	0.47	0.05	
7	0.00	0.68	0.32	0.01	
8	0.00	0.82	0.18	0.01	
9	0.00	0.95	0.05	0.03	
10	0.92	0.00	0.08	0.01	
11	0.83	0.00	0.17	0.03	
12	0.72	0.03	0.25	0.04	
13	0.57	0.12	0.32	0.05	
14	0.51	0.23	0.27	0.02	
15	0.34	0.38	0.28	0.04	
16	0.29	0.50	0.21	0.01	
17	0.15	0.69	0.16	0.04	
18	0.02	0.92	0.06	0.07	
19	0.16	0.84	0.00	0.03	
20	0.35	0.65	0.00	0.03	

(d) 예측 혼합비율 (잡음 수준:0.1)

표본	S_1 비율	S_2 비율	S_3 비율	\sum	오차
1	0.00	0.00	1.00	0.07	
2	0.00	0.11	0.89	0.06	
3	0.00	0.23	0.77	0.05	
4	0.00	0.31	0.69	0.06	
5	0.00	0.40	0.60	0.07	
6	0.00	0.59	0.41	0.01	
7	0.00	0.70	0.30	0.00	
8	0.00	0.83	0.17	0.02	
9	0.00	0.93	0.07	0.02	
10	0.91	0.03	0.07	0.01	
11	0.82	0.00	0.18	0.03	
12	0.76	0.01	0.24	0.06	
13	0.54	0.13	0.32	0.06	
14	0.46	0.24	0.30	0.03	
15	0.34	0.31	0.35	0.08	
16	0.26	0.51	0.24	0.03	
17	0.21	0.66	0.14	0.01	
18	0.00	0.92	0.08	0.07	
19	0.20	0.80	0.00	0.00	
20	0.43	0.57	0.00	0.02	

표 4.3: 잡음 수준에 따른 예측정확도의 비교

잡음 수준	평균절대오차	평균제곱오차
0.01	0.03	0.0405
0.05	0.04	0.0455
0.1	0.04	0.0518

표 4.3의 요약표를 보면 잡음수준 0.01, 0.05, 0.1에 대한 평균절대오차는 각각 0.03, 0.04, 0.04로서 잡음수준이 높아져도 구성비에 대한 예측정확도는 크게 영향을 받지 않음을 볼 수 있다. 이는 ICA에 의한 독립구성성분 도출과정에서 잡음의 영향이 어느정도 제거되기 때문이라고 보인다. 따라서 독립성분분석을 이용한 혼합물의 미지성분 비율 예측은 분광 데이터에서 가질 수 있는 잡음을 고려하더라도 그 성분비율을 예측하는데 안정적인 방법이라고 하겠다.

6. 결론

ICA는 오디오 프로세스, 의학영상 시그널분석, 이미지 분석, 통신분야, 계량경제분야 등 다양한 분야에서 그 적용범위가 넓어지고 있는 통계적 기법으로서, 독립성분이 미지인 경우 혹은 성분을 계량화하여 이차적 분석이 필요한 경우 유용한 다변량기법이다.

본 연구에서는 혼합물의 미지 구성성분을 분리하고 분리된 성분을 이용하여 성분간 비율을 예측하는 방안을 제시하고, 도금강판의 엑스선 회절강도 데이터에 적용하여 비율을 예측하여 품질지표로 사용할 수 있는 사례를 분석하였다. 또한 ICA에 의한 성분도출과 NNLS를 이용한 비율이 적합하게 추정되는지 평가하기 위하여 모의실험을 수행한 결과 여러가지 잡음수준에서도 안정적인 예측이 가능함을 보였다.

참고문헌

- Hyvärinen, A. (1999). "Survey on independent component analysis", *Neural Computing Surveys*, 2, 94-128.
- Hyvärinen, A. and Oja, E. (2000). "Independent component analysis: algorithms and applications", *Neural Networks*, 13(4-5), 411-430.
- Ikeda, S. and Toyama, K. (2000). "Independent component analysis for noisy data-MEG data analysis", *Neural Networks* 13, 1063-1074.
- Lawson, C. L. and Hanson, R. J. (1974). *Solving least squares problems*, Prentice-Hall, Englewood Cliffs, NJ, Chapter 23
- Lee, D. D. and Seung, H. S. (1999). "Learning the parts of objects by non-negative matrix factorization", *Nature*, vol. 401, 788-791.
- Lee, D. D. and Seung, H. S. (2001). "Algorithms for nonnegative matrix factorization," in *Advances in Neural Information Processing Systems 13*. Cambridge, MA: MIT Press, 556-562

Vigário, R., Jousmäki, V., Hyvärinen, M., Hari, R. and Oja, E. (1998). Independent component analysis for identification of artifacts in magnetoencephalographic recordings, *Advances in Neural Information Processing Systems*, Vol. 10. Cambridge, MA:MIT Press (pp. 229-235).

[2005년 7월 접수, 2005년 11월 채택]

Predicting Unknown Composition of a Mixture Using Independent Component Analysis

Hyeseon Lee¹⁾ Jaekee Song²⁾ Haesang Park³⁾ Chi-Hyuck Jun⁴⁾

ABSTRACT

Independent component analysis (ICA) is a statistical method for transforming an observed high-dimensional multivariate data into statistically independent components. ICA has been applied increasingly in wide fields of spectrum application since ICA is able to extract unknown components of a mixture from spectra. We focus on application of ICA for separating independent sources and predicting each composition using extracted components. The theory of ICA is introduced and an application to a metal surface spectra data will be described, where subsequent analysis using non-negative least square method is performed to predict composition ratio of each sample. Furthermore, some simulation experiments are performed to demonstrate the performance of the proposed approach.

Keywords: Independent component analysis, Non-negative least square, non-Gaussian

-
- 1) Graduate Student, Department of Statistics, Kyungpook National University, 370 Sankyuk-dong, Buk-gu, Daegu 702-701, Korea
E-mail : hyelee@postech.ac.kr
 - 2) Professor, Department of Statistics, Kyungpook National University, 370 Sankyuk-dong, Buk-gu, Daegu 702-701, Korea
E-mail:jksong@knu.ac.kr
 - 3) Graduate student, Department of Industrial & Management Engineering, POSTECH, San 31 Hyoja-dong, Pohang 790-784, KOREA
E-mail : shoo359@postech.ac.kr
 - 4) Professor, Department of Industrial & Management Engineering, POSTECH, San 31 Hyoja-dong Pohang 790-784, KOREA
E-mail : chjun@postech.ac.kr