

신경망을 이용한 로버스트 주성분 분석에 관한 연구¹

김상민² 오광식³ 박희주⁴

요약 주성분 분석은 자료압축, 특징추출, 통신이론, 패턴인식 그리고 화상처리등의 컴퓨터 공학분야에서 중요하게 사용되고 있다. Oja(1982,1989,1992)는 확률적 경사 강하법(SGA:Stochastic Gradient Ascent)을 이용한 제한된 헵규칙을 제안하여 주성분 분석에 사용하였다. 그러나, 이 규칙은 이상치에 민감하므로 이상치의 영향을 줄이기 위해, Xu & Yuille(1995)는 통계물리 방법을 이용한 로버스트 에너지함수를 생성하여 로버스트 주성분 분석방법을 제안하였다. 또한 Devlin et.al(1981)은 M-추정량을 이용하여 주성분 분석을 하였다. 본 논문에서는 Oja(1992)의 규칙과 Xu & Yuille(1995)의 로버스트 에너지함수를 이용하여 신경망을 구성하였다. 그리고, Devlin et.al(1981)이 제안한 시뮬레이션조건하에서 실험을 하였다. 실험한 결과와 Devlin et.al(1981)의 결과를 비교, 분석함으로써, 신경망의 성능을 확인하고자 한다.

주제어: 로버스트 주성분분석, 신경망, 헵규칙

1. 서론

신경망은 컴퓨터 공학 분야에서 폭 넓고 다양한 과학적 문제들을 해결하기 위해 활발히 연구되어 각 분야에 응용되어져 왔다. 최근에는 통계분야에서 신경망을 어떻게 활용할 수 있는지에 대한 연구가, 일반적인 통계문제에 적용시키기 위해서(Ripley(1994) and Ripley(1996)) 또는 회귀 모형에 대한 문제에 적용시키기 위해서(Hwang et.al(1994)) 그리고 주성분 분석을 하기 위해서(Xu & Yuille(1995)) 진행되어져 왔다. 특히, 주성분 분석은 자료압축, 특징추출, 통신이론, 패턴인식 그리고 화상처리등의 컴퓨터 공학분야에서 중요하게 사용되고 있다. 주성분 분석을 위한 방법으로는 기존의 통계적인 방법 뿐만 아니라, 투사지향(Projection Pursuit)방법 및 신경망(Neural Networks)방법등이 있다. 본 논문에서는 신경망을 이용하여 주성분을 구하는 방법에 대해 연구하고자 한다.

Oja(1992)는 헵 규칙에 제한조건을 추가하여 제한된 헵규칙(Constrained Hebbian Learning Rule)을 제안하였다. 확률적 경사강하법(SGA:Stochastic Gradient Ascent)을 이용한 이 규칙은 가장 큰 고유치에 대응되는 고유벡터부터 차례대로 구해낸다. 특히, 입력 변수가 많고 몇개의 고유벡터만을 구하여 주성분 분석을 할 때에는 이 방법이 유용하게 사용되어 질 수 있다.

그러나 이 규칙은 이상치에 민감하기 때문에, Xu & Yuille(1995)는 통계물리 방법을 이용한 로

1 본 연구는 대구효성가톨릭대학교의 연구비에 의하여 연구 되었음.

2 경북 김천시 삼락동 김천전문대학 전산정보처리과, 740-200

3 경북 경산시 하양읍 대구효성가톨릭대학교 통계학과, 712-702

4 경북 경산시 하양읍 경북산업대학교 전자계산학과, 712-701

버스트 에너지함수를 구하여 이상치에 민감하지 않은 규칙인 로버스트 주성분 분석(RPCA:Robust Principal Component Analysis)방법을 제안하였다. 또한 Devlin et. al (1981)은 M-추정량을 이용하여 이상치의 영향을 최소화 시켜 주성분 분석을 하였다.

이에 본 논문에서는 Oja(1992)의 제한된 헵규칙과, Xu & Yuille(1995)의 로버스트 에너지 함수를 이용하여 이상치의 영향을 최소화시켜 고유벡터를 추정하도록 신경망을 구성하여 시뮬레이션을 하였다. 표본의 추출 조건으로는, Devlin et.al(1981)이 사용한 시뮬레이션 조건과 동일한 조건인 정규분포, 오염된 대칭정규분포, 오염된 비대칭정규분포, 그리고 코시분포 대신 자유도 5인 Elliptical t-분포를 이용하였다. 실험한 시뮬레이션의 결과와 Devlin et.al(1981)의 M-추정량으로 구한 결과를 비교, 분석함으로써, 신경망의 성능을 확인하고자 한다.

2. 로버스트 주성분 분석

2.1 Oja의 주성분 분석 방법

Oja(1992)는 확률적 경사 강하법을 이용하여, 제한된 헵규칙을 다음과 같이 제안하였고, 이 규칙은 제일 큰 고유치부터, 차례대로, 대응되는 고유벡터 전부를 구해낸다. 아래 가중치벡터가 수렴 후에 고유벡터가 된다.

$$\underline{w}_j(t+1) = \underline{w}_j(t) + \alpha(t)y_j(t)(\underline{x}_i(t) - y_j(t)\underline{w}_j(t) - \sum_{k < j} y_k(t)\underline{w}_k(t))$$

단 $\underline{w}_j(t)$: 주어진 시점 t 에서의 j -번째 가중치벡터

$\underline{x}_i(t)$: i -번째 관측치 벡터

$y_j(t) = \underline{w}_j(t)^T \underline{x}_i(t)$: 주어진 시점 t 에서의 j -번째 출력값.

2.2 Xu & Yuille의 로버스트 주성분 분석 방법

그러나 위의 규칙은 이상치에 민감하기 때문에, 이상치의 영향력을 감소시키기 위하여, Xu & Yuille(1995)는 통계물리 방법을 이용하여 로버스트 에너지 함수를 구하였으며, 이 로버스트 에너지함수를 추가한 로버스트 주성분 분석규칙을 다음과 같이 작성할 수 있다.

$$\underline{w}_j(t+1) = \underline{w}_j(t) + \alpha(t) \frac{(\underline{x}_i(t) - y_j(t)\underline{w}_j(t) - \sum_{k < j} y_k(t)\underline{w}_k(t))}{1 + \exp(\beta(z(\underline{x}_i(t), \underline{w}_j(t)) - \eta))}$$

$$\text{단 } z(\underline{x}_i(t), \underline{w}_j(t)) = \left(\underline{x}_i(t)^T \underline{x}_i(t) - \frac{\underline{w}_j(t)^T \underline{x}_i(t) \underline{x}_i(t)^T \underline{w}_j(t)}{\underline{w}_j(t)^T \underline{w}_j(t)} \right),$$

β 와 η 는 임의의 상수.

2.3 Devlin et.al.의 주성분 분석 방법

Devlin et.al(1981)은 M-추정량을 사용하여 이상치의 영향을 최소화 시켜 상관행렬을 추정하였고, 고유치와 고유벡터를 추정하여 주성분 분석을 하였다. 표본으로는 4가지 분포인 정규분포, 오염된 대칭정규분포, 오염된 비대칭정규분포, 코시분포를 이용하였다. 앞의 4가지 분포로부터

50행에 6열의 자료를 만들어 6개의 고유벡터를 추정하였으며, 이 과정을 200회 반복 실시하였다.

3. 모의실험

위의 로버스트 주성분 분석 규칙을 이용하여 은닉층이 없는 6개의 입력노드와 6개의 출력노드로 구성된 신경망을 다음 그림 1과 같이 구성하여 시뮬레이션을 진행하였다.

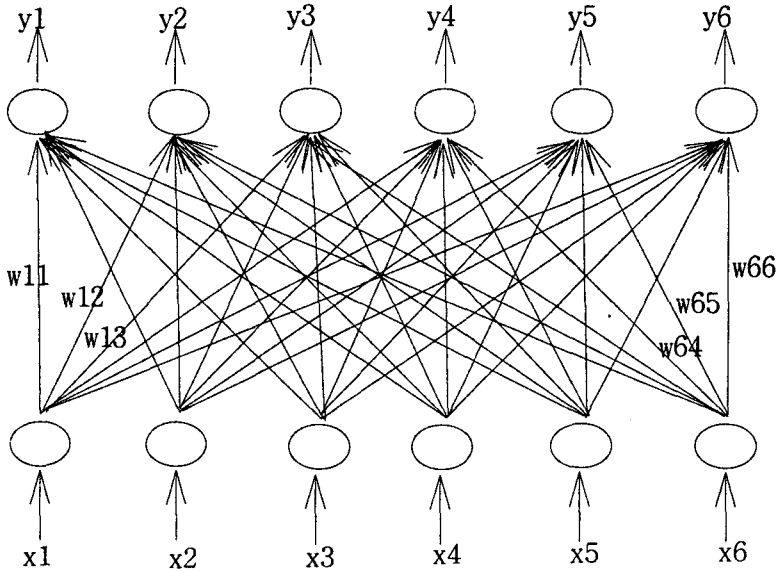


그림 1. 은닉층이 없는 6개의 입력노드와 6개의 출력노드로 구성된 신경망

시뮬레이션에 사용한 자료는 Devlin et.al(1981)과 Li & Chen(1985)이 사용한 조건으로 실행하였다. 즉 6차원의 4가지 분포(정규분포, 오염된 대칭분포, 오염된 비대칭분포, 코시분포 대신 자유도 5인 Elliptical t-분포)에 대한 상관행렬 P 의 고유벡터를 추정하는 문제를 고려하는 것이다. 이때 상관행렬 P , 고유치 λ , 목표치 고유벡터 α , 및 4 가지 분포는 다음과 같이 주어진다.

$$P = \begin{bmatrix} P_1 & 0 \\ 0 & P_2 \end{bmatrix}$$

$$\text{단 } P_1 = \begin{bmatrix} 1 & 0.95 & 0.30 \\ 0.95 & 1 & 0.10 \\ 0.30 & 0.10 & 1 \end{bmatrix}, \quad P_2 = \begin{bmatrix} 1 & -0.499 & -0.499 \\ -0.499 & 1 & -0.499 \\ -0.499 & -0.499 & 1 \end{bmatrix}.$$

고유치 : $\lambda = \text{diag}(2.029, 1.499, 1.499, 0.943, 0.028, 0.002)$

고유벡터 : $\alpha'_6 = (1/\sqrt{3})(0, 0, 0, 1, 1, 1)$ 를 사용하였다.

분포: 가) Normal: $NOR[\underline{0}, P]$

나) Symmetric contaminated normal: $SCN[\underline{0}, P]$ with $.9NOR[\underline{0}, P] + .1NOR[\underline{0}, P]$

다) Elliptical t-distribution on 5 degrees of freedom : $ET[\underline{0}, P]$

라) Asymmetric contaminated normal: $ACN[\underline{\mu}, P]$

with $.9NOR[\underline{0}, P] + .1NOR[\underline{\mu}, P]$ where $\underline{\mu} = .537\alpha_6$

위와 같은 조건하에서 4가지 각 분포에 대해 50행의 6열 관측치를 생성하고, 시뮬레이션을 실시하여 고유벡터를 추정하였다. 추정된 고유벡터는 6개로 이는 가중치 벡터의 수와 같다. 이와 같은 과정을 200회 실행하여 고유벡터를 추정하였다.

4. 토의

본 논문에서는 Oja(1992)의 제한된 행규칙과, Xu & Yuille(1995)의 로버스트 에너지 함수를 이용하여 이상치의 영향을 최소화시켜 고유벡터를 추정하도록 신경망을 구성하여 시뮬레이션을 하였다. 실험한 결과와 Devlin et.al(1981)의 M-추정량을 사용하여 구한 결과를 비교, 분석함으로써, 신경망의 성능을 확인하고자 한다.

목표치인 모상관행렬의 고유벡터와 Xu & Yuille(1995)의 로버스트 주성분 분석 방법을 사용하여 추정한 고유벡터 간의 근접성(closeness)을 측정하기 위해 $|\cos \theta_i|$ 를 사용하였다. 여기서 $\hat{\theta}_i$ 은 두 고유벡터 간의 각을 나타낸다.

목표치로 세운 고유벡터와 실험을 통해 추정한 고유벡터와의 근접성 결과를 그림 2에 첨부하였다. 세로로는 차례로 $NOR(0, P)$, $SCN(0, P)$, $ACN(\mu, P)$, $ET(0, P)$ 를 가로로는 4개의 고유벡터에 대해 200회의 근접성 값을 히스토그램으로 보였다. 목표치로 세운 고유벡터와 실험을 통해 추정한 고유벡터와의 근접성이 고유벡터에 따라 차이를 보이고 있다.

그림 2의 실험 결과를 보면 4가지 분포 모두에서, 첫번째, 네번째에서 추정한 고유벡터의 근접성이 목표치와 가까운 것으로 나타났다. Devlin et.al(1981)에서는 정규분포에 대해서만 첫번째 고유벡터가 잘 추정되었다. 그러나, M-추정량을 적용하였을 때는 다섯번째와 여섯번째 고유벡터

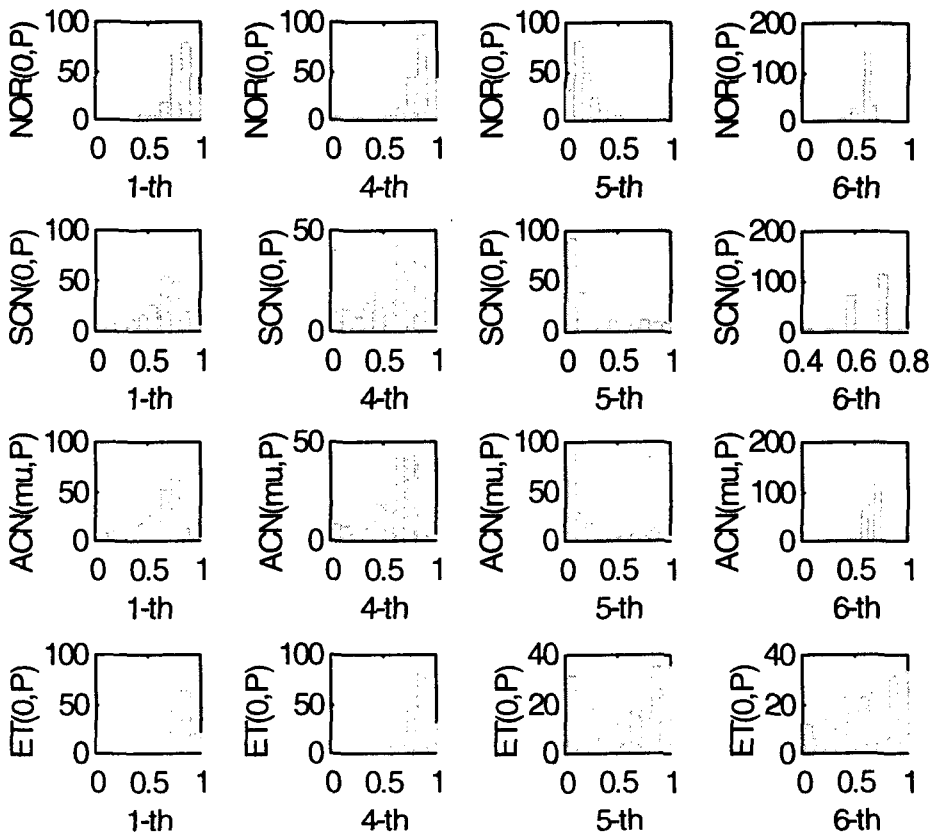


그림 2. 시뮬레이션 결과의 히스토그램

가 잘 추정되었다. 그러나 본 논문에서의 시뮬레이션 결과에서 추정된 고유벡터는 4가지 분포 모두에서 첫번째, 네번째 고유벡터를 잘 추정하였다. 이것은 각 분포 모두에서 이상치의 영향을 최소화 시켰기 때문이다.

한편 다섯번째, 여섯번째 고유벡터를 잘 추정하지 못한것은 고유치가 너무 작기 때문이다. 또한 두번째, 세번째 고유벡터에 대한 근접성은 고유치 $\lambda_2 = \lambda_3$ 가 동일하기 때문에, 각도 θ_i 는 α_2 와 α_3 에 의해 생성(span)된 공간으로 투사된 것과 $\hat{\alpha}_i$ 사이의 값이 되어 비교할 의미가 없으므로 그림 2에서 삭제되었다.

따라서 신경망으로 구한 로버스트 주성분 분석 방법의 강점은, 이상치의 효과를 최소로 줄이면서, 특히 입력변수가 많고 몇 개의 고유벡터를 구할 경우에는 기존의 통계적 방법보다 이 방법이 좋은 결과를 나타낼 수 있다. 본 논문의 로버스트 주성분 분석 방법을 이용하여 대단위의 입력 자료에 대하여 단지 몇 개의 주성분을 분석하여 최소한의 정보를 보관함으로써 자료압축, 화상처리, 정보통신 등의 정보처리 분야에 적용할 수 있다. 본 신경망 프로그램은 C 언어로 586PC에서 구현되었다.

참고문헌

- Devlin, S. J., Gnanadesikan, R. and Kettenring, J. R. (1981). *Robust Estimation of Dispersion Matrices and Principal Components*, Journal of the American Statistical Association, **76**, 354-362.
- Hwang, J., Lay, S., Maechier, M., Martin, R. D., and Schiment, J. (1994), *Regression modeling in back-propagation and projection pursuit learning*. IEEE Transactions on neural networks, **5**, 342-353.
- Li, G. and Chen, Z. (1985), *Projection-Pursuit Approach to Robust Dispersion Matrices and Principal Components: Primary Theory and Monte Carlo*. Journal of the American Statistical Association, **80**, 759-766.
- Oja, E. (1982), *A Simplified Neuron Model as a Principal Component Analyzer*. Journal of Mathematical Biology, **15**, 267-278
- Oja, E. (1989), *Neural Networks, Principal Components, and Subspaces*. International Journal of Neural Systems, **1**, 61-68.
- Oja, E. (1992), *Principal Components, Minor Components, and Linear Neural Networks*, Neural Networks, **5**, 927-935.
- Ripley, B. D. (1994), *Neural networks and related methods for classification (with discussion)*, Journal of the Royal Statistical Society Series B, **56**, 409-456.
- Ripley, B. D. (1996). *Pattern Recognition and Neural Networks*. Cambridge: Cambridge University Press.
- Softky, W. R. and Kammen, D. M. (1991), *Correlations in High Dimensional or Asymmetric Data Sets: Hebbian Neuronal Processing*. Neural Networks, **4**, 337-347.
- Xu, L. and Yuille, A. L. (1995), *Robust Principal Component Analysis by Self-Organizing Rules Based on Statistical Physics Approach*. IEEE trans on Neural Networks, **6**, 131-143.

On Robust Principal Component using Analysis Neural Networks

Sangmin Kim⁵, Kwangsik Oh⁶ and Heejoo Park⁷

Abstract Principal component analysis(PCA) is an essential technique for data compression and feature extraction, and has been widely used in statistical data analysis, communication theory, pattern recognition, and image processing. Oja(1992) found that a linear neuron with constrained Hebbian learning rule can extract the principal component by using stochastic gradient ascent method.

In practice real data often contain some outliers. These outliers will significantly deteriorate the performances of the PCA algorithms. In order to make PCA robust, Xu & Yuille(1995) applied statistical physics to the problem of robust principal component analysis(RPCA). Devlin et.al(1981) obtained principal components by using techniques such as M-estimation.

The propose of this paper is to investigate from the statistical point of view how Xu & Yuille's(1995) RPCA works under the same simulation condition as in Devlin et.al(1981).

Keywords: Hebbian rule, Neural Networks, Robust Principal Component Analysis.

⁵ Dept. of Computer Information Processing, Kimchun Junior College, Kyungbuk 740-200

⁶ Dept. of Statistics, Catholic University of Taegu-Hyosung, Kyungbuk 712-702

⁷ Dept. of Computer Science, Kyungbook Sanup University, Kyungbuk 712-701