비선형 특징 추출을 위한 온라인 비선형 주성분분석 기법

(On-line Nonlinear Principal Component Analysis for Nonlinear Feature Extraction)

김 병 주 † 심 주 용 †† 황 창 하 ††† 김 일 곤 ††††
(Byung-Joo Kim) (Joo-Yong Shim) (Chang-Ha Hwang) (Il-Kon Kim)

요 약 본 논문에서는 온라인 학습 자료의 비선형 특징(feature) 추출을 위한 새로운 온라인 비선형 주성분분석(OL-NPCA : On-line Nonlinear Principal Component Analysis) 기법을 제안한다. 비선형 특징 추출을 위한 대표적인 방법으로 커널 주성분방법(Kernel PCA)이 사용되고 있는데 기존의 커널 주성분 분석 방법은 다음과 같은 단점이 있다. 첫째 커널 주성분 분석 방법을 N 개의 학습 자료에 적용할 때 $N \times N$ 크기의 커널 행렬의 저장 및 고유벡터를 계산 하여야 하는데, N의 크기가 큰 경우에는 수행에 문제가 된다. 두 번째 문제는 새로운 학습 자료의 추가에 의한 고유공간을 새로 계산해야 하는 단점이 있다. OL-NPCA는 이러한 문제점들을 점진적인 고유공간 갱신 기법과 특징 사상 함수에 의해 해결하였다. Toy 데이타와 대용량 데이타에 대한 실험을 통해 OL-NPCA는 다음과 같은 장점을 나타낸다. 첫째 메모리 요구량에 있어 기존의 커널 주성분분석 방법에 비해 상당히 효율적이다. 두 번째 수행 성능에 있어 커널 주성분 분석과 유사한 성능을 나타내었다. 또한 제안된 OL-NPCA 방법은 재학습에 의해 쉽게 성능이 향상 되는 장점을 가지고 있다.

키워드 : 온라인 비선형 주성분분석, 커널 주성분분석, 특징 사상 함수, 고유값, 고유벡터

Abstract The purpose of this study is to propose a new on-line nonlinear PCA(OL-NPCA) method for a nonlinear feature extraction from the incremental data. Kernel PCA(KPCA) is widely used for nonlinear feature extraction, however, it has been pointed out that KPCA has the following problems. First, applying KPCA to N patterns requires storing and finding the eigenvectors of a $N \times N$ kernel matrix, which is infeasible for a large number of data N. Second problem is that in order to update the eigenvectors with an another data, the whole eigenspace should be recomputed. OL-NPCA overcomes these problems by incremental eigenspace update method with a feature mapping function. According to the experimental results, which comes from applying OL-NPCA to a toy and a large data problem, OL-NPCA shows following advantages. First, OL-NPCA is more efficient in memory requirement than KPCA. Second advantage is that OL-NPCA is comparable in performance to KPCA. Furthermore, performance of OL-NPCA can be easily improved by re-learning the data.

Key words : On-line nonlinear Principal Component Analysis, Kernel Principal Component Analysis, Nonlinear feature extractor, Eigenvalue, Eigenvector

1. 서 론

주성분분석(Principal Component Analysis:PCA) 방법은 학습 자료의 특징추출을 위해 사용하는 대표적인 기법중의 하나이다[1]. 하지만 PCA 방법은 다음과 같은 단점이 있다. 첫째 PCA는 일괄처리(batch) 방식으로 동작한다. 이는 새로운 학습 자료가 추가 되면 고유공간(eigenspace)을 다시 계산 하여야 하는 단점이 있다. 두 번째 문제점은 PCA의 적용범위는 학습자료 간의 선형 상관관계(linear correlation)가 있는 경우로만 국한된다.

이러한 문제점을 해결하기 위한 기존의 연구를 살펴

† 정 회 원 : 영산대학교 네트워크정보공학부 교수
bjkim@ysu.ac.kr
†† 비 회 원 : 대구가톨릭대학교 정보통계학과 교수
ds1631@hanmail.net
††† 비 회 원 : 대구가톨릭대학교 정보통계학과 교수
chhwang@cataegu.ac.k
†††† 종신회원 : 경북대학교 컴퓨터과학과 교수
ikkim@knu.ac.kr
논문접수 : 2003년 1월 10일
심사완료 : 2003년 12월 9일

보면 다음과 같다. 첫째 일괄처리 방식의 문제점을 해결하기 위해 학습 자료를 순차적으로 받아들이며 이전의 고유공간과 새로이 추가된 학습 자료에 의해 새로운 고유공간을 계산하는 온라인 방식은 크게 두 가지로 나눌 수 있다. 하나는 새로운 고유공간을 계산하는데 있어 평균의 갱신을 허용하지 않는 방법이며[2,3] 또 다른 접근 방법은 Hall[4]에 의해 제안된 방법으로 평균의 갱신을 허용하는 방법이다. Hall은 그의 논문에서 평균을 갱신하는 방법이 고정하는 방법에 비해 분류문제에서 우수한 성능을 나타냄을 실험을 통해 보였다.

비선형 PCA에 관한 기존의 연구를 살펴보면 다음과 같다. Tipping[5]은 선형 주성분분석 기법을 결합한 MPCA(mixture PCA)방법으로 비선형 문제를 해결하고자 하였다. 또 다른 접근방법으로 복원오차(reconstruction error)를 최소화 하는 자동연상(autoassociative) 다층 퍼셉트론 등이 제안되었다[6,7]. 이러한 해결책들은 모두 비선형 최적화 기법을 요구할 뿐 아니라 목적함수(objective function)가 종종 지역 최소화에 빠질 가능성이 있다. 최근에 제안된 방법 중 Scholkopf[8]는 커널함수(kernel function)를 이용한 커널 주성분 분석(KPCA) 방법을 제안하였다. 이는 선형 주성분분석 방법에서와 유사한 방법으로 고유치(eigenvalue) 문제를 해결함으로써 비선형 특징을 추출한다.

비록 KPCA 방법이 비선형 자료에 대해 특징 추출을 수행할 수 있지만 다음과 같은 단점이 있다. KPCA는 학습 자료의 개수가 N 개 일 때 $N \times N$ 크기의 커널 행렬을 저장해야 할 뿐만 아니라 커널행렬의 고유값을 구해야 한다. 만약 N 의 값이 큰 경우에는 고유값을 계산하기 위해 많은 양의 메모리가 필요하다.

최근에 학습 자료가 큰 경우에 대해 KPCA의 문제점을 해결하는 몇몇 시도가 있었다. Rosipal[9]은 계산상에 있어 단순하고 효율적인 EM(Expectation Maximization)알고리즘을 KPCA에 적용하는 방법을 제안하였다. 하지만 이 방법도 일괄처리 방식으로 수행되어 커널 행렬을 저장해야 하는 문제점을 해결 하지는 못했다. Moerland[10]는 EM 알고리즘을 KPCA에 적용하는 온라인 커널 주성분 분석 방법을 제안 하였다. Scholkopf[11]에 의해 제안된 축약된 집합 선택 기법(reduced set selection method)은 KPCA의 메모리 문제를 어느 정도 해결 하였으나 많은 계산량이 요구 되는 단점이 있다. Smola[12]는 희소 커널 특징 분석(sparse kernel feature analysis)을 제안하여 KPCA의 계산상의 문제점을 해결하였다. 하지만 이 방법은 커널 행렬을 저장해야 하는 문제점을 해결하지는 못했다.

본 논문에서는 앞에서 열거한 기존의 KPCA의 단점을 해결하기 위해 Hall에 의해 제한된 고유공간 갱신 기법과 특징사상 함수를 사용한 온라인 비선형 PCA (OL-NPCA) 방법을 제안한다. 논문의 구성은 다음과 같다. 2장에서는 온라인 PCA 기법에 대해 간략히 설명한다. 3장에서는 특징사상 함수를 사용한 OL-NPCA 알고리즘을 설명한다. 4장에서는 OL-NPCA 방법의 타당성 및 효율성을 검정하기 위해 실험에 의해 나타난 결과를 보인다. 마지막 5장에서는 결론 및 향후 연구에 대해서 이야기 한다.

2. 온라인 PCA

Hall에 의해 제안된 온라인 PCA 방법을 간략히 설명하기 전에 먼저 수식에서 사용되는 벡터 및 행렬은 다음과 같이 정의한다. 벡터는 열벡터(column vector)를 의미하며 소문자로 나타내고 행렬은 대문자로 표시한다. 그리고 행렬의 크기는 아랫첨자로 나타낸다. 예를 들어 A_{mn}는 $m \times n$ 크기의 행렬을 의미한다. 행렬에서 열의 확장(column extension)은 대괄호(square brackets)로 나타낸다. 따라서 $[A_{mn} \quad b]$는 $m \times (n+1)$ 크기의 행렬을 나타내며 벡터 b는 행렬 A의 마지막 열에 추가된다.

동적인 고유공간 갱신 기법을 설명하기 위해 다음을 정의한다. $U_{nk} = [u_j], j = 1 \cdots k$는 현재까지 학습 자료 $x_i, i = 1 \cdots N$에서 구한 고유벡터 집합을 나타내며, $\lambda = diag(\Lambda_{nn})$ 는 고유치 행렬(eigenvalue matrix) Λ_{nn}의 대각요소를 내림차순으로 정렬한 것을 나타내며 $\bar{x}$는 x 의 평균을 의미한다.

온라인 PCA 방법은 학습자료 x_{N+1}이 추가 되었을 때 이전 학습 자료의 저장 없이 갱신하는 것이다. 먼저 새로운 학습 자료가 추가 되었을 때 갱신된 평균은 식 (1)과 같이 구한다.

$$\bar{x}' = \frac{1}{N+1}(N\bar{x} + x_{N+1}) \tag{1}$$

식 (1)에 의해 새로운 평균이 구해지면 추가된 학습 자료에 의해 갱신된 고유 벡터 집합을 구할 수 있다. 이를 위해서는 이전의 고유벡터를 회전행렬(rotational matrix)에 적용하여야 하는데 이를 위해 먼저 직교잔차 벡터(orthogonal residual vector)를 구해야 한다. 직교잔차벡터는 식 (2)와 같이 계산된다.

$$\hat{h} = (U_{nk} a_{N+1} + \bar{x}) - x_{N+1} \tag{2}$$

식 (2)에서 구한 $\hat{h}$을 정규화한 것은 식 (3)과 같이 표시된다.

$$h_{N+1} = \frac{h_{N+1}}{\|h_{N+1}\|_2} \text{ for } \|h_{N+1}\|_2 \text{ and } h_{N+1} = 0 \text{ otherwise} \tag{3}$$

여기서 $\|h_{N+1}\|_2$은 노름(norm)을 의미한다. 노름은 절대값을 일반화한 개념으로 벡터의 크기를 재는 척도

이며 $|x|_2=\sqrt{x_1^2+\ldots+x_n^2}$ 으로 표현한다. 예를 들어 $x=(1,2,-3,2)^T$일 때 $|x|_2=\sqrt{1+4+9+4}=3\sqrt{2}$와 같이 계산된다.

$q=k+1$ 라 정의하면 새로운 고유벡터 U'_{nq}은 식 (3)에 의해 구해진 h_{N+1}과 이전의 고유벡터 U_{nk}에 의해 생성된 행렬을 회전행렬 $R_{(k+1)(k+1)}$에 적용하여 구할 수 있으며 식 (4)와 같이 구한다.

$$U'_{nq}=[U_{nk}\ \ h'_{N+1}]R_{(k+1)(k+1)} \tag{4}$$

여기서 $R_{(k+1)(k+1)}$은 회전행렬 이며 식 (5)의 고유공간의 해이다.

$$D_{(k+1)(k+1)}R_{(k+1)(k+1)}=R_{(k+1)(k+1)}\Lambda'_{(k+1)(k+1)} \tag{5}$$

여기서 $\Lambda'_{(k+1)(k+1)}$은 새로운 고유치들의 대각 행렬이다. 행렬 $D_{(k+1)(k+1)}$는 다음과 같이 구성할 수 있다.

$$D_{(k+1)(k+1)}=\frac{N}{N+1}\begin{bmatrix}\Lambda_{kk} & 0\\ 0^T & 0\end{bmatrix}+\frac{N}{(N+1)^2}\begin{bmatrix}aa^T & \gamma a\\ \gamma a^T & \gamma^2\end{bmatrix} \tag{6}$$

여기서 $\gamma=h^T_{N+1}(x_{N+1}-\bar{x})$, $a=U^T(x_{N+1}-\bar{x})$와 같이 구하며 0는 k차원의 영벡터(zero vector)이다. 행렬 $D_{(k+1)(k+1)}$를 구성하는 몇 가지 방법이 제안되었는데 Hall이 제안한 방법만이 평균을 갱신할 수 있도록 제공하는데 이 기법은 평균의 갱신을 허용하지 않는 기법에 비해 성능이 우수한 것으로 알려져 있다[4]. 따라서 OL-NPCA 방법에서는 Hall이 제안한 평균의 갱신을 허용하는 방법을 이용한다.

2.1 고유공간 갱신 기준

온라인 PCA 방법을 적용하면 순차적인 학습 자료에 대해 k 개의 고유벡터가 만들어 지며 학습 자료는 주성분(principal component)을 이용하여 나타낼 수 있다. 주성분에 의해 표현되는 학습 자료는 식 (7)과 같이 나타낸다.

$$\hat{x}_{i(N)}=U_{nk}a_{i(N)}+\bar{x} \tag{7}$$

새로운 학습자료 x_{N+1}에 의해 주성분을 갱신하기 위해 식 (8)과 같이 보조벡터(auxiliary vector) η를 정의한다.

$$\eta=[U_{nk}\hat{h}_{N+1}]^T(\bar{x}-\bar{x}') \tag{8}$$

식 (8)에서 구한 η를 이용하여 모든 학습 자료에 대한 주성분은 다음과 같이 구한다.

$$a_{i(N+1)}=(R'_{(k+1)(k+1)})^T\begin{bmatrix}a_{i(N)}\\ 0\end{bmatrix}+\eta,\quad i=1\cdots N+1 \tag{9}$$

식 (9)에 의해 이미 학습되어진 N 개의 자료와 새로 추가된 학습자료 x_{N+1}을 완전히 재구성(reconstruction) 할 수 있다. 하지만 이러한 재구성은 차원이 하나 늘어남으로 인해 재구성 오차는 줄일 수 있는 가능성은 높아지지만 더 많은 기억공간을 필요로 하는 문제점을 가지고 있다. 따라서 한편으로는 고유벡터 공간 k의 차원이 증가 하는 것을 막아야 하며 또 다른 한편으로는 식 (9)에 의해 재구성 되는 학습 자료의 오차를 줄이기 위해 고유벡터 공간 k의 차원을 늘여야 하는데 이 두 가지 조건을 적절히 만족하는 임계점을 찾는 것이 필요하다. 주성분의 개수를 어느 정도까지 유지해야 하는지에 대한 명확한 기준은 없다. 본 논문에서는 Joliffe[1]에 의해 제안된 방법을 사용하며 식 (10)과 같다.

$$(N+1)\lambda'_{k+1}>0.7\bar{\lambda} \tag{10}$$

여기서 $\bar{\lambda}$은 λ의 평균이다. 이 규칙에 의해 새로운 고유벡터 u'_{k+1}의 추가 여부를 결정한다.

3. 온라인 비선형 PCA(OL-NPCA)

2장에서 설명한 온라인 PCA 방식은 그 적용 범위가 학습 자료간의 선형적인 관계가 존재할 때 적용이 가능하다. 학습 자료 간에 비선형성이 존재하는 경우에 대해 온라인 PCA를 적용하고자 할 경우 Cover[13]의 정리가 하나의 해결책이 될 수 있다. 그의 정리를 간단히 요약하면 "몇 가지 조건이 만족될 때 입력 공간에서 선형적으로 분리 불가능한 문제가 비선형 사상을 적용한 고차원 특징 공간(feature space)에서는 선형적으로 분리 가능한 문제로 변환될 확률이 높다"이다. 그림 1은 이러한 예를 보여주고 있다. 비록 고차원 공간으로의 사상이 학습 자료간의 비선형 문제를 해결할 수는 있지만 고차원 공간으로의 사상으로 인한 계산 비용의 증가를 초래한다. 이러한 문제는 커널 기법(kernel trick)으로 해결이 가능하다. 커널 함수 $K(x_i, x_j)$는 다음과 같이 표현할 수 있다.

$$K(x_i,\ x_j)\ =\ (\phi(x_i)\cdot\phi(x_j)) \tag{11}$$

식 (11)이 의미하는 것은 만약 특징 사상 ϕ를 적절히 선택하면 특징 공간에서의 내적(inner product)은 입력 공간에서의 커널과 동일하다는 것으로 고차원의 특징 공간에서 계산할 필요가 없어 계산상의 증가를 줄일 수 있다는 것이다. 커널 함수는 Mercer[14]의 조건을 만족하는 함수들이 사용 가능하다고 알려져 있으며 그러한 함수는 표 1과 같은 함수이다.

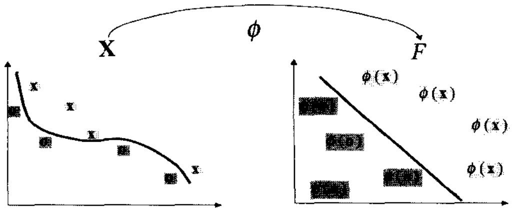


그림 1 입력 공간의 특징 공간으로의 사상

표 1 대표적인 커널함수

커널함수	종류
$(x\cdot y)^d$, $((x\cdot y)+1)^d$	d차 다항식
$\exp(-\frac{(x-y)^2}{2\sigma^2})$	가우시안
$\tanh((x\cdot y)-offset)$	다층 퍼셉트론

커널함수 K가 양반정치행렬(semi positive definite)이면 특징 사상 함수 ϕ가 존재한다는 것을 Vapnik[15]은 증명하였다. 하지만 다항식 커널 함수를 제외한 나머지 커널 함수에서 특징 사상 함수 ϕ를 구하는 것은 상당히 어려운 문제이다. 다항식 커널함수에서 특징 사상 함수 ϕ를 구하는 과정은 다음과 같다. 설명을 위해 d=2, $x=(x_1, x_2)$, $y=(y_1, y_2)$라 두면,

$$(x\cdot y)^2=(x_1^2, \sqrt{2}x_1x_2, x_2^2)(y_1^2, \sqrt{2}y_1y_2, y_2^2)^T = (\phi(x)\cdot\phi(y))$$

따라서 $\phi(x)=(x_1^2, \sqrt{2}x_1x_2, x_2^2)$이 된다. 다항식 특징 사상 함수의 경우 차원 d 의 값에 따라 성능은 큰 차이를 보이지 않는다[16].

따라서 본 논문에서는 KPCA의 일괄처리 방식의 문제점은 Hall이 제안한 온라인 PCA기법을 적용하여 해결하고, 선형 자료에만 적용이 가능한 문제는 다항식 커널 함수에서 구해진 특징 사상함수를 온라인 PCA에 적용하여 비선형이면서 온라인 학습 자료에 대해서도 적용할 수 있는 온라인 비선형 주성분분석 알고리즘을 제안한다.

4. 실험 및 결과

일반적으로 온라인 방법은 일괄처리 방식에 비해 메모리 효율성이 좋고 학습 자료의 추가를 허용하는 유연한 기법이지만 일괄처리 방식에 비해 정확도가 떨어지는 것이 문제가 된다. 따라서 실험은 먼저 toy 자료에 대해 OL-NPCA와 KPCA와의 주성분값, 재구성 오차(reconstruction error), 고유벡터의 비교 등을 통해 제안된 OL-NPCA의 정확도를 검증한다. 이때 사용하는 데이타는 Scholkopf[7]가 그의 논문에서 사용한 비선형 자료를 사용한다. Toy 데이타에 대해 OL-NPCA의 타당성을 검정한 후 대용량 데이타에 대한 OL-NPCA의 메모리 효율성, 정확도 및 학습의 유연함을 검증하기 위해 기계학습 분야에서 많이 사용하는 바나나 학습 자료 및 UCI machine learning repository 자료에 대해 적용한다.

4.1 Toy 자료

Toy 자료는 다음과 같이 생성한다. 여기서 ε은 평균 0, 분산 1인 정규분포로부터 생성된 임의의 오차항이다.

$$y=x^2+0.2\varepsilon \quad : \quad \varepsilon\sim N(0,1), x=[-1,1] \tag{12}$$

이때 사용하는 다항식 특징 사상 함수의 차원 d는 2로 하였다.

식 (12)에 의해 생성된 자료에 대해 KPCA와 OL-NPCA를 적용한 결과가 그림 2에 나타나 있다. 그림 2의 왼쪽은 제안된 OL-NPCA 방법이며 오른쪽은 일괄처리 KPCA 방법이다. 위에서부터 아래 방향으로 고유값을 내림차순으로 정렬하여 처음 3개의 값을 나타내었다. 이때 등고선은 같은 주성분 값들을 연결한 것이다. 그림에서 알 수 있듯이 두 방법에서 구해진 고유치와 주성분 그래프는 매우 유사한 형태를 나타낸다.

OL-NPCA의 성능 평가는 재구성오차(reconstruction error), 그리고 KPCA와 OL-NPCA에서 구해진 고유벡터의 cos θ 값에 의한 유사도에 의해서도 평가할 수 있다. 먼저 재구성 오차는 다음과 같이 정의한다.

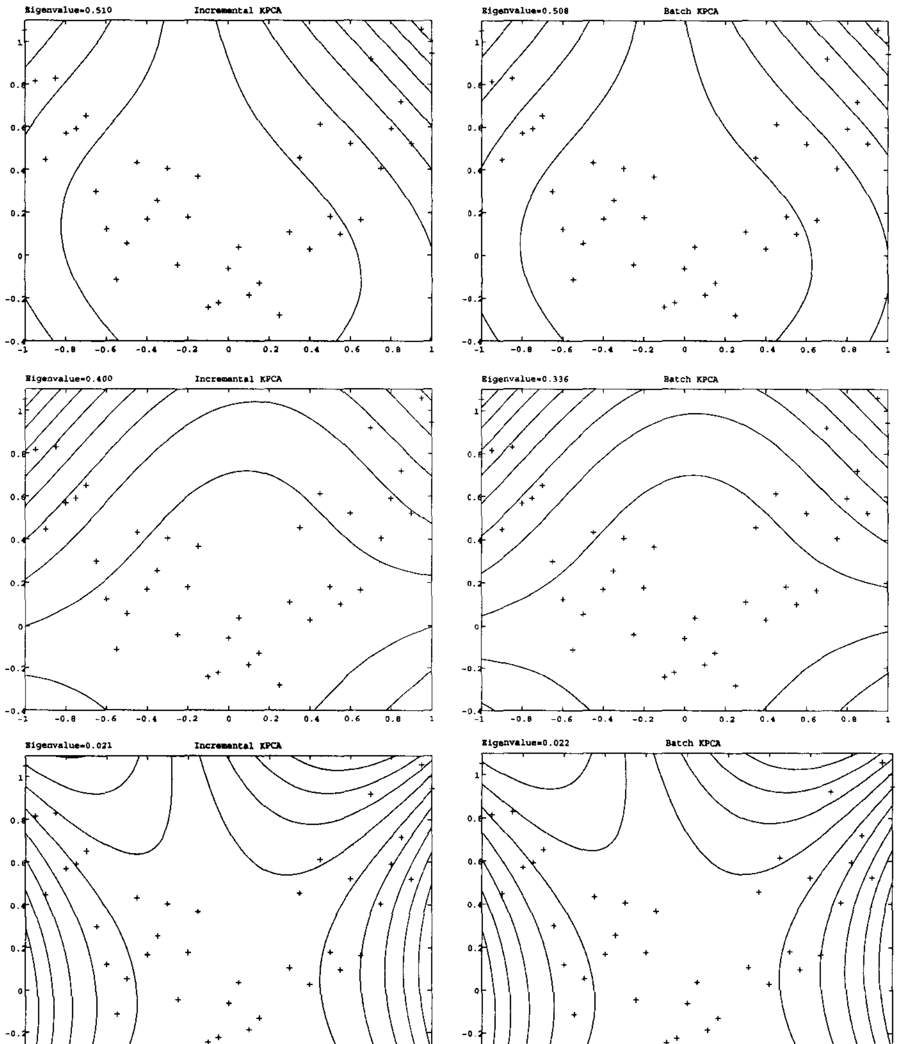
$$\delta=|\phi(x^n)-P_l\phi(x^n)|^2 \tag{13}$$

여기서 P_l는 OL-NPCA 방법에 의해서 구해진 주성분이다. 그림 3은 식 (13)에서 정의한 재구성 오차를 재학습에 의해 제곱 평균 오차(Mean Square Error: MSE)가 감소하는 과정을 나타낸 그래프이다. 순차적으로 입력된 학습 자료에 대해 제안된 방법에 의한 재구성 오차의 MSE는 0.0090이다. MSE의 값이 0.0090이라는 의미는 특징공간에서 학습 데이타와 OL-NPCA에 의해서 추정된 학습 데이타의 오차가 거의 없다는 것을 의미한다. 이는 제안된 OL-NPCA 방법이 일괄처리 방법과 유사한 성능을 나타냄을 보여준다.

OL-NPCA 방법의 또 다른 장점은 재학습(re-learning)에 의해 성능을 향상 시킬 수 있다는 것이다. 그림 3에서 보면 재학습 회수가 증가함에 따라 재구성 오차가 감소함을 볼 수 있다. 재구성 오차는 재학습이 진행되면서 감소하다 학습 횟수 30정도에서 오차의 감소가 거의 없는 수렴 상태가 된다. 또 다른 성능평가 방법으로 일괄처리 KPCA와 OL-NPCA 에서 구한 고유벡터가 이루는 cos θ 값의 유사도에 의해 평가할 수 있다. 아래의 행렬 B와 I는 각각 일괄처리 커널주성분 방법과 OL-NPCA을 실험에 적용하여 구한 고유벡터 행렬이다.

$$B=\begin{bmatrix} 0.53856 & -0.095116 & 0.81604 \\ 0.094315 & 0.49981 & -0.19613 \\ 0.82981 & -0.11002 & -0.5259 \\ 0.11171 & 0.85384 & 0.13795 \end{bmatrix}$$

$$I=\begin{bmatrix} 0.53856 & -0.095114 & -0.8166 \\ 0.094314 & 0.49981 & 0.19364 \\ 0.82981 & -0.11002 & 0.52635 \\ 0.11171 & 0.85384 & -0.13649 \end{bmatrix}$$



(a) 일괄처리 커널주성분방법 (b) OL-NPCA 방법

그림 2 일괄처리 커널 주성분 방법과 제안된 OL-NPCA 방법

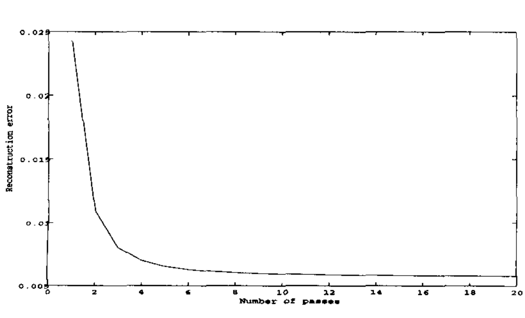


그림 3 OL-NPCA방식에서 재학습에 의한 재구성오차의 변화

두 행렬의 고유벡터가 이루는 $\cos\theta$는 다음과 같이 정의 한다.

$$\cos\theta = \frac{\alpha \cdot \beta}{|\alpha|\,|\beta|} \tag{14}$$

식 (14)에 의해 구해진 $\cos\theta$와 θ값이 표 2에 나타나 있다. 표에서 알 수 있듯이 모든 고유벡터의 θ 값은 0 이다. 즉 두 벡터가 같다는 것을 의미한다. 고유벡터를 비교한 결과에서도 일괄처리 커널주성분 방법과 OL-NPCA 성능이 같음을 알 수 있다.

표 2 일괄처리 주성분방법과 OL-NPCA 방법에서 구한 고유벡터의 cos와 값

고유벡터	$\|\cos\theta\|$	θ
1	1	0
2	1	0
3	1	0

4.2 대용량 자료

제안된 OL-NPCA 방법의 메모리 효율성 및 정확도를 검증하기 위해 대용량의 학습 자료에 대한 실험을 수행하였다. 학습 자료는 기계학습에서 학습 알고리즘의 성능을 평가 하는 벤치마킹 자료로 많이 사용되는 바나나 학습 데이타이다(http://www.first.gmd.de/~raetsch). 바나나 학습 자료는 2개의 클래스로 구성된 비선형 자료이며 학습 알고리즘의 성능을 평가하기 위해 학습용 자료 400개와 테스트 자료 4900개로 분리되어 있다. 하지만 본 논문에서는 분류가 아닌 특징 추출이 목적이므로 두개를 합친 5300개의 학습 자료에 대해 실험을 하였다. 학습 자료의 분포는 그림 4에 나타나 있다. 그림 3에서 두개의 클래스는 'x'와 '.'로 표시되어 있다. 이때 사용하는 다항식 특징 사상 함수의 차원 d는 2로 하였다.

바나나 학습 자료에 대한 OL-NPCA의 수행결과 5번의 재학습에 의한 MSE는 0.0021이다. 앞의 학습 결과와 마찬 가지로 대용량의 학습 데이타에 대해서도 OL-NPCA의 수행 능력이 KPCA와 유사함을 알 수 있다. 반면에 메모리 효율 면에서 보면 바나나 학습 자료를 KPCA에 적용할 경우 고유벡터를 구하기 위해 5300 × 5300 크기의 커널 행렬을 저장하여야 한다. 하지만 OL-NPCA는 커널 행렬을 저장할 필요 없이 4 × 4 크기의 회전행렬 R 과 4 × 4 크기의 D 행렬 정도만이 필요하다.

표 3에 KPCA와 OL-KPCA와의 메모리 요구량을 비교하였다. 표 3에서 메모리 효율성이란 OL-KPCA가 필요로 하는 메모리를 1로 설정하였을 때 KPCA를 수행하기 위해 필요한 메모리의 상대적인 값이다. 표에서 알 수 있듯이 학습 자료의 개수가 클수록 OL-NPCA이 매우 유용한 기법임을 보여주고 있다. 바나나 학습 자료에 대해서도 OL-NPCA는 재학습에 의해 성능이 개선되는 것을 장점으로 보여주고 있으며 이는 그림 5에 나타나 있다.

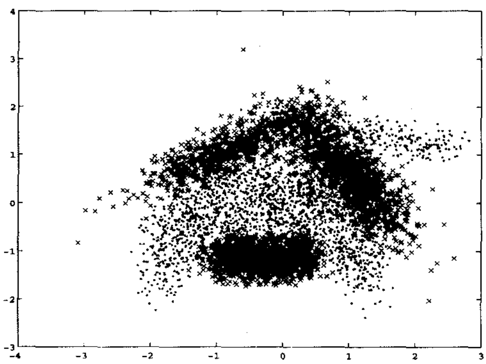


그림 4 바나나 학습 자료의 분포

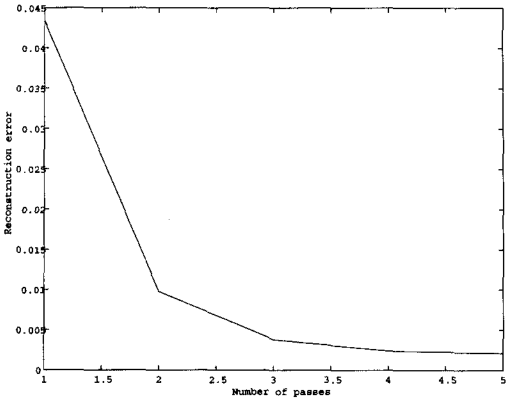


그림 5 OL-NPCA를 바나나 자료에 적용시 재학습에 의한 재구성오차의 변화

표 3 바나나자료에 대한 KPCA와 OL-KPCA의 메모리 요구량 비교

	KPCA	OL-KPCA
Kernel 행렬	$O(5300\times5300)$	필요없슴
R 행렬	필요없슴	$O(4\times4)$
D 행렬	필요없슴	$O(4\times4)$
효율성(KPCA / IKPCA)	2,247,200	1

4.3 현실세계 자료

제안된 방법의 현실문제에의 적용을 위해 UCI machine learning repository 중에서 갑상선자료에 대해 실험하였다. 갑상선 자료는 972개의 자료로 구성되어 있으며 각 학습 자료는 29개의 속성으로 구성되어 있다. 학습 자료는 3개의 그룹으로 분류되어 있으며 자료에 대한 자세한 설명은 http://ftp.ics.uci.edu/pub/machine-learning-databases/thyroid-disease/를 참조하면 된다. 실험에서 다항식 특징 사상 함수의 차원 d는 앞의 실험과 같이 2로 하였다. 갑상선 학습 자료의 경우 차원이 29인데 학습 자료의 차원이 3 이상인 경우 n을 학습 자료의 차원이라고 하면 사상함수 $\phi(x)$는 식 (15)와 같이 일반화시킬 수 있다.

$$\phi(x) = (x_1^2, \ldots, x_n^2, \sqrt{2}x_1x_2, \ldots \sqrt{2}x_{n-1}x_n) \qquad (15)$$

갑상선 자료의 경우 식 (15)에 의해 특징공간으로 사

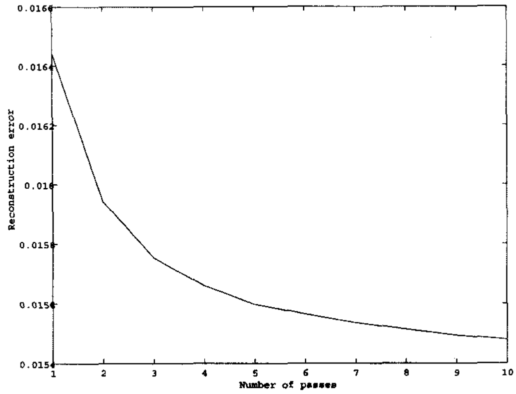


그림 6 OL-NPCA를 갑상선 자료에 적용시 재학습에 의한 재구성오차의 변화

상된 각 학습자료의 차원은 406이 된다. 갑상선 자료에 대한 OL-NPCA의 수행결과가 그림 6에 나타나 있다. 10번의 재학습에 의한 재구성 오차의 MSE는 0.0155이다. 앞의 학습 결과와 마찬 가지로 현실세계의 학습 데이타에 대해서도 OL-NPCA의 수행 능력이 KPCA와 유사함을 알 수 있다.

반면에 메모리 효율 면에서 보면 갑상선 학습 자료를 KPCA에 적용하여 고유벡터를 계산하기 위해 972 × 972 크기의 커널 행렬을 저장하여야 한다. 하지만 OL-NPCA는 커널 행렬을 저장할 필요 없이 91 × 91 크기의 회전행렬 R과 91 × 91 크기의 D 행렬 정도만이 필요하다. 표 4에 각 방법에 대한 메모리 요구량을 나타내었다. 갑상선 학습 자료에 대해서도 OL-NPCA는 재학습에 의해 성능이 개선되는 것을 그림 6을 통해 알 수 있다.

표 4 갑상선자료에 대한 KPCA와 OL-KPCA의 메모리 요구량 비교

	KPCA	OL-NPCA
Kernel 행렬	$O(972\times972)$	필요없슴
R행렬	필요없슴	$O(91\times91)$
D 행렬	필요없슴	$O(91\times91)$
효율성(KPCA/OL-NPCA)	57.0453	1

5. 결 론

본 논문에서는 비선형 특징(feature) 추출을 위한 새로운 OL-NPCA 기법을 제안한다. 제안된 OL-NPCA 기법은 다음과 같은 의미를 가진다.

첫째 비선형 자료의 특징 추출 성능에서는 OL-NPCA 방법이 기존의 일괄처리 KPCA와 유사한 성능을 나타내었다.

둘째 기존의 비선형 추출 기법에 비해 메모리 사용면에서 효율적이다. 기존의 커널 주성분 분석방법의 경우 학습 자료의 개수가 N개 일 때 고유공간(eigenspace)을 계산하기 위해 $O(N^2)$ 만큼의 메모리가 필요하다. 반면에 제안된 OL-NPCA의 경우 k를 학습중에 유지해야할 고유벡터 행렬의 갯수라고 하면 $O((k+1)^2)$ 만큼의 메모리가 필요하다. 이는 학습 자료의 개수 N이 클 경우에는 OL-NPCA이 매우 유용하다.

세 번째 학습 자료의 추가에 의한 유연성이다. 기존의 일괄처리 방식의 경우 새로운 학습 자료가 추가 되면 고유벡터와 고유값을 새로 계산해야 하는 단점이 있다. 하지만 제안된 OL-NPCA 방법은 학습 자료의 추가를 허용 하면서 학습 및 비선형 특징을 추출할 수 있다.

앞으로의 연구 방향은 OL-NPCA의 재학습 시 학습을 중단하는 시점을 결정하는 문제이다. 본 논문에서는 일정 수만큼 재학습을 하였지만 신경망의 학습에서와 같이 학습을 중단하는 규칙이 필요하다.

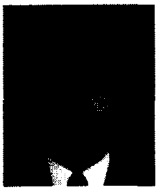
참 고 문 헌

[1] I.T. Jolliffe, "Principal Component Analysis," New York Springer-Verlag, 1986.
[2] H. Murakami, B.V.K. V Kumar., "Efficient calculation of primary images from a set of images," IEEE PAMI, 4(5), pp.511-515, 1982.
[3] J. Winkeler, B.S. Manjunath and S. Chandrasekaran., "Subset selection for active object recognition," In CVPR, volume 2, pp.511-516, IEEE Computer Society Press, June 1999.
[4] P. Hall, D. Marshall, and R. Martin., "Incremental eigenalysis for classification," In British Machine Vision Conference, volume 1, pp. 286-295, September 1998.
[5] M.E. Tipping and C.M. Bishop., "Mixtures of probabilistic principal component analysers," Neural Computation 11(2), pp.443-482, 1998.
[6] M.A. Kramer., "Nonlinear principal component analysis using autoassociative neural networks," AICHE Journal 37(2), pp.233-243, 1991.
[7] K.I. Diamantaras and S.Y. Kung., Principal Component Neural Networks: Theory and Applications, New York John Wiley & Sons, Inc. 1996.
[8] B. Scholkopf, A. Smola and K.R. Muller., "Nonlinear component analysis as a kernel eigenvalue problem," Neural Computation 10(5), pp.1299-1319, 1998.
[9] R. Rosipal and M. Girolami., "An Expectation Maximization approach to nonlinear component analysis," Neural Computation, 13(3) pp.505-510, 2001.
[10] P. Moerland, "An on-line EM algorithm applied to

kernel PCA," IDIAP Research Report. 2000.
[11] B. Scholkopf, S. Mika, C. Burges, P. Knirsch, K.R. Miller, G. Ratsch and A.J. Smola., "Input Space versus Feature Space in Kernel-based Methods," IEEE Transactions on Neural Networks, vol. 10, pp.1000-1017, September 1999.
[12] A.J. Smola, O.L. Mangasarian, and B. Scholkopf., "Sparse kernel feature analysis," Technical Report 99-03, University of Wisconsin, Data Mining Institute, Madison, 1999.
[13] Cover, "Geometrical and statistical properties of system of linear inequalities with applications in pattern recognition," IEEE Transactions on Electronic Computers, Vol. EC-14, 326-334. 1965.
[14] J. Mercer., "Functions of positive and negative type and their connecrion with the theory of integral equations," Philos. Trans. Roy. Soc. London, Vol. 209, pp.415-446, 1909.
[15] V. N. Vapnik., Statistical learning theory. John Wiley & Sons, New York, 1998.
[16] P. Moerland, "Mixture models for unsupervised and supervised learning," Ph.D. thesis, IDIAP research report IDIAP-PR 00-18, 2000.

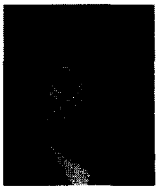
김 병 주

1990년 부산대학교 전자계산학과 졸업(이학사). 1992년 부산대학교 대학원 전자계산학 전공(이학석사). 2003년 경북대학교 대학원 컴퓨터과학 전공 박사(이학박사). 1992년~1996년 서라벌 대학 전자계산과 교수. 1996년~2002년 성심외국어대학 정보통신학부 교수. 2003년~현재 영산대학교 네트워크정보공학부 교수. 관심분야는 기계학습(machine learning)



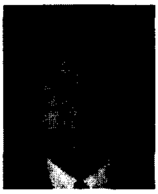
심 주 용

1983년 서강대학교 화학공학과 졸업. 1988년 Northeastern Illinois University 대학원 수학과 졸업(이학석사). 1992년 University of Illinois at Urbana-Champaign 대학원 통계학과 박사 수료 1996년 경북대학교 대학원 통계학과(이학박사). 2003년~현재 대구가톨릭대학교 정보통계학과 겸임교수. 관심분야는 dynamic programming, machine learning



황 창 하

1982년 경북대학교 사범대학 수학교육과(이학사). 1984년 서울대학교 대학원 계산통계학과(이학석사). 1985년~1987년 한국통신 전임연구원. 1991년 미국 미시간대학교 통계학과(이학박사). 1992년~1994년 경성대학교 전산통계학과 조교수 1995년~현재 대구가톨릭대학교 정보통계학과 부교수. 관심분야는 기계학습, 정보과학



김 일 곤

1991년 서울대학교 전산과학과 박사학위 취득. 1992년~현재 경북대학교 컴퓨터과학과 교수로 근무. 1997년 3월~1998년 8월 미국 죠지타운대학교 병원 ISIS 연구소 방문연구자. 관심분야는 지능형 에이전트 시스템, 분산시스템, 의료정보학