
잡음 민감성이 개선된 퍼지 주성분 분석

허경용* · 우영운** · 김성훈***

An Improved Robust Fuzzy Principal Component Analysis

Gyeongyong Heo* · Young Woon Woo** · Seong Hoon Kim***

요 약

주성분 분석(PCA)은 데이터의 차원을 줄이면서 최대의 데이터 변이를 보존하는 기법으로 차원 축소나 피쳐 추출을 위해 널리 사용되고 있다. 하지만 PCA는 잡음에 민감한 단점이 있으며, 이러한 잡음 민감성을 해결하기 위해 여러 가지 PCA 변형이 제안되었다. 그 중 robust fuzzy PCA(RF-PCA)는 퍼지 소속도를 사용하여 잡음의 영향을 효과적으로 줄일 수 있음이 입증되었다. 하지만 RF-PCA 역시 몇 가지 문제점이 있고, 수렴성이 그 중 하나이다. RF-PCA는 소속도와 주성분을 갱신할 때 서로 다른 목적 함수를 사용하므로 수렴 속도가 느리고 구해지는 해가 국부 최적 해임을 보장하지 않는다. 이 논문에서는 RF-PCA의 문제점을 해결하기 위해 하나의 목적 함수를 이용해 소속도와 주성분을 갱신할 수 있는 방법을 제안한다. 제안한 방법, RF-PCA2는 반복 최적화를 이용함으로써 국부 최적해에 수렴함을 보장하며, RF-PCA에 비해 빠른 수렴 속도를 가지고, 잡음 민감성이 줄어든다. 이러한 사실들은 실험 결과를 통해 확인할 수 있다.

ABSTRACT

Principal component analysis (PCA) is a well-known method for dimension reduction while maintaining most of the variation in data. Although PCA has been applied to many areas successfully, it is sensitive to outliers. Several variants of PCA have been proposed to resolve the problem and, among the variants, robust fuzzy PCA (RF-PCA) demonstrated promising results. RF-PCA uses fuzzy memberships to reduce the noise sensitivity. However, there are also problems in RF-PCA and the convergence property is one of them. RF-PCA uses two different objective functions to update memberships and principal components, which is the main reason of the lack of convergence property. The difference between two functions also slows the convergence and deteriorates the solutions of RF-PCA. In this paper, a variant of RF-PCA, called RF-PCA2, is proposed. RF-PCA2 uses an integrated objective function both for memberships and principal components. By using alternating optimization, RF-PCA2 is guaranteed to converge on a local optimum. Furthermore, RF-PCA2 converges faster than RF-PCA and the solutions found are more similar to the desired solutions than those of RF-PCA. Experimental results also support this.

키워드

주성분 분석, 퍼지 소속도, 수렴성, 잡음 민감성, Robust 퍼지 PCA

Key word

Principal Component Analysis, Fuzzy Membership, Convergence Property, Noise Sensitivity, Robust Fuzzy PCA

* Computer and Information Science and Engineering, University of Florida (제1저자)

** 동의대학교 멀티미디어공학과

*** 경북대학교 컴퓨터정보학부 (교신저자, shkim1454@knu.ac.kr)

접수일자 : 2010. 01. 26

심사완료일자 : 2010. 02. 12

I. 서 론

주성분 분석(principal component analysis, PCA)은 차원 축소를 위해 널리 사용되는 방법 중 하나로, 자승 오차의 합 (sum squared error) 또는 재건 에러(reconstruction error)를 최소화하는 직교 변환이다[1]. PCA는 여러 영역에 성공적으로 사용되었지만, 잡음에 민감하다는 문제점이 있으며, 이러한 잡음 민감성은 재건 에러 최소화라는 PCA의 목적에서 연유한다. PCA의 잡음 민감성을 줄이기 위해 여러 가지 방법이 제안되었고, 이들은 부분 집합을 이용하는 방법과 퍼지 소속도(membership)을 이용하는 방법의 두 가지로 크게 나누어볼 수 있다. 부분 집합을 이용하는 방법들은 하나 이상의 데이터 부분 집합에 대해 PCA를 수행하고 이들 결과를 이용하여 전체 데이터의 주성분을 구해낸다[2][3][4]. 이들 방법이 PCA의 잡음 민감성을 줄여주는 하지만, 데이터 포인트의 수가 적은 경우에는 사용하기 어려운 단점이 있다. 퍼지 소속도를 이용하는 방법은 전형적인 데이터 포인트에만 큰 소속도 값을 할당함으로써 잡음 민감성을 줄인다. 하지만 이들 방법의 문제점은 소속도를 구해야 한다는 점이다. 퍼지 소속도를 이용하는 대부분의 방법은 먼저 소속도를 구하고, 구해진 소속도를 이용하여 퍼지 공분산 행렬로부터 주성분을 구하는 두 단계로 이루어진다[5]. 소속도 기반의 방법들은 소속도를 구하는 첫 번째 단계에 초점을 두며[6][7], 이 과정은 목적 함수를 최적화하는 과정으로 나타내어진다. 하지만 이들 방법은 목적 함수가 첫 번째 주성분에만 의존하여 결정된다는 문제점이 있다. 또 다른 문제점은 소속도를 구하기 위해 사용하는 목적 함수가 PCA와 연관성이 없다는 점이다.

잡음 민감성을 개선하기 위해 제안된 여러 방법 중 robust fuzzy PCA(RF-PCA)는 k 개의 주성분을 동시에 고려하고, 소속도를 계산하는 과정에서 재건 에러를 최소화하는 반복 최적화(iterative optimization) 방법을 사용하였다[5]. PCA 역시 재건 에러를 최소화하므로, 소속도를 계산할 때 재건 에러를 최소화하는 것은 자연스러운 일이다. 하지만 RF-PCA는 퍼지 소속도를 사용하는 방법의 두 단계인 소속도와 주성분 계산 과정에서 서로 다른 목적 함수를 사용함으로써 알고리즘이 수렴함을 보장하지 못하는 문제점이 있다. 비록 RF-PCA의 수

렴성이 실험적으로 보여 졌지만[5], 소속도의 최대 차이 값이 0으로 수렴한다는 것만이 보여졌다. 또한 두 가지 목적 함수를 사용하는 것은 RF-PCA의 수렴 속도를 느리게 할뿐만 아니라 구해진 해가 잡음의 영향을 받도록 한다.

이 논문에서는 RF-PCA의 문제점을 해결하기 위해 향상된 RF-PCA, RF-PCA2를 제안한다. RF-PCA2는 소속도와 주성분을 구하기 위해 하나의 목적 함수를 사용하고 이를 반복 최적화 기법으로 최적화함으로써 구해진 해가 국부 최적해임을 보장한다[8]. 또한 RF-PCA2는 RF-PCA에 비해 수렴 속도가 빠르고 잡음 민감성이 줄어드는 것을 실험 결과를 통해 확인할 수 있다.

다음 장에서는 RF-PCA2를 위한 목적 함수를 제안하고 소속도와 주성분의 update equation을 구한다. 또한 RF-PCA의 목적 함수와 비교함으로써 그 차이를 보인다. 3장에서는 실험을 통해 RF-PCA와 RF-PCA2를 비교한다. 결론 및 향후 연구 방향은 4장에서 언급한다.

II. 향상된 Robust Fuzzy PCA

이 장에서는 소속도와 주성분을 갱신하기 위해 공통의 목적 함수를 사용하는 향상된 RF-PCA, RF-PCA2를 제안한다. RF-PCA2의 목적 함수는 식 (1)과 같이 나타낼 수 있다.

$$\begin{aligned} \operatorname{argmin}_{W} J = & \sum_{i=1}^N u_i \| (x_i - \mu_R) - WW^T(x_i - \mu_R) \|^2 \quad (1) \\ & + \sigma^2 \sum_{i=1}^N (u_i \log u_i - u_i) \\ = & \sum_{i=1}^N u_i e(x_i) + \sigma^2 \sum_{i=1}^N (u_i \log u_i - u_i) \\ \text{s.t.} \quad & W^T W = I \end{aligned}$$

이 때 N 은 데이터 포인트의 개수를, u_i 는 데이터 포인트 x_i 의 소속도를, μ_R 은 가중치 평균을, $e(x_i)$ 는 재건 에러[9]를 나타내고, W 는 k 개 단위 직교 벡터를 열로 가지는 행렬을 나타내며 σ 는 regularization 상수이다. k 는 재건에 사용될 직교 벡터의 개수를 나타내며 알려진 것으로 가정한다. 식 (1)은 재건 에러의 가중치 합

을 최소로 하는 k 개 직교 벡터를 구하기 위한 목적 함수로 이들 직교 벡터가 잡음에 민감하지 않은 고유 벡터에 해당한다. 식 (1)은 두 개의 항으로 구성되어 있다. 첫 번째 항은 소속도를 가중치로 가지는 재건 에러의 합을 나타내고 두 번째 항은 주성분의 잡음 민감성을 줄이기 위한 regularization 항이다. Regularization 항 역시 두 개의 항으로 구성된다. 첫 번째 항은 음의 엔트로피로 퍼지 클러스터링에서 잡음 민감성을 줄이기 위해 널리 사용되는 항이다[10]. 퍼지 클러스터링에서 엔트로피를 regularization 항으로 사용하는 경우 목적함수는 식 (2)와 같다.

$$J_{Entropy} = \sum_{k=1}^M \sum_{i=1}^N u_{ki} d^2(x_i, v_k) + \sigma^2 \sum_{k=1}^M \sum_{i=1}^N u_{ki} \log u_{ki} \quad (2)$$

이 때 M 는 클러스터의 개수를, u_{ki} 는 x_i 가 k 번째 클러스터에 소속되는 정도를 나타내며, $d(x_i, v_k)$ 는 x_i 와 클러스터 중심 v_k 사이의 거리를 나타낸다. 퍼지 클러스터링에서는 M 개의 클러스터에 대한 소속도의 합이 1이 되는 제약 조건이 있다. 이 조건은 regularization 항과 더불어 잡음 민감성을 줄이는 역할을 한다. 하지만 RF-PCA2에서는 하나의 클러스터만 존재하며 이러한 제약 조건이 없다. 따라서 소속도 항을 첨가함으로써 엔트로피 항이 자명해(trivial solution)를 갖지 않도록 해준다.

식 (1)을 u_i 에 대해 편미분하면 식 (3)의 update equation을 얻을 수 있다.

$$u_i = \exp\left(-\frac{e(x_i)}{\sigma^2}\right) \quad (3)$$

또한 식 (1)을 μ_R 에 대해 편미분하면 가중치 평균에 대한 update equation을 식 (4)와 같이 구할 수 있다.

$$\mu_R = \frac{\sum_{i=1}^N u_i x_i}{\sum_{i=1}^N u_i} \quad (4)$$

벡터 μ_R 은 소속도를 가중치로 가지므로 가중치 평균

이라 불린다. 가중치 평균을 이용하여 주어진 데이터 집합은 영의 가중치 평균을 가지도록 만들 수 있고, $x'_i = x_i - \mu_R$, 식 (1)을 최적화하는 W 를 구하는 것은 식 (5)를 최적화하는 것과 같아진다.

$$J' = \sum_{i=1}^N u_i \|x'_i - WW^T x'_i\|^2 \quad (5)$$

$W^T W = I$ 의 제약 조건 하에서 식 (5)를 최소화하는 W 는 식 (6)으로 정의되는 가중치 분산 행렬(weighted covariance matrix)에서 k 개의 최대 고유치를 가지는 고유 벡터로 구성할 수 있다. 이에 관한 증명은 부록에 첨부하였다.

$$C_R = \frac{1}{N} \sum_{i=1}^N u_i x'_i x'^T_i \quad (9)$$

$$= \frac{1}{N} \sum_{i=1}^N u_i (x_i - \mu_R)(x_i - \mu_R)^T$$

1. 소속도 $U_0 = [u_1, \dots, u_N]^T = [1, \dots, 1]^T$ 와 카운터 $t = 0$ 를 초기화하고 W_0 를 PCA를 사용하여 초기화한다.
2. repeat
3. $t \leftarrow t + 1$
4. 평균 μ_R 과 분산 행렬 C_R 을 구한다.
5. C_R 의 고유 벡터로 W_t 를 구성한다.
6. 식 (5)로 소속도 U_t 를 구한다.
7. until $J^{t-1} - J^t < \epsilon$ 또는 $t > t_{\max}$
8. return

그림 1. RF-PCA2 알고리즘
Fig. 1. RF-PCA2 algorithm

RF-PCA2 알고리즘은 그림 1과 같이 요약할 수 있으며, 이 때 ϵ 은 미리 정해지는 상수이고 t_{\max} 는 최대 반복 횟수를 나타낸다.

RF-PCA2에 사용되는 가중치 평균과 가중치 분산 행렬은 robust 통계의 M-estimator와 관련이 있다[11]. Robust 통계에서 평균과 분산에 대한 M-estimator는 식

(7) 및 (8)로 정의된다.

$$\mu_M = \frac{\sum_{i=1}^N f_1(g(x_i))x_i}{\sum_{i=1}^N f_1(g(x_i))} \quad (7)$$

$$C_M = \frac{1}{M} \sum_{i=1}^N f_2(g^2(x_i))(x_i - \mu_M)(x_i - \mu_M)^T \quad (8)$$

이 때 f_1 과 f_2 는 단조 감소 함수이고 g 는 거리 척도를 나타낸다. f_1 과 f_2 를 재전 에러와 지수함수를 이용하여 식 (9) 및 (10)과 같이 정의하면 식 (4) 및 (6)은 식 (7) 및 (8)과 동일해진다.

$$f_1(x) = \exp\left(-\frac{e(x)}{\sigma^2}\right) = u_i \quad (9)$$

$$f_2(x) = \left(\left(\exp\left(-\frac{e(x)}{\sigma^2}\right)\right)^2\right)^{1/2} = u_i \quad (10)$$

즉, 가중치 평균과 가중치 분산은 평균과 분산에 대한 robust estimator임을 알 수 있다.

RF-PCA에서 역시 식 (3)이 소속도를 갱신하기 위해 사용된다. 즉, RF-PCA에서 소속도를 구하기 위해서는 식 (1)을 최적화한다. 하지만 RF-PCA에서 가중치 분산 행렬은 식 (6)과는 다르게 식 (11)로 정의된다.

$$C' = \frac{1}{N} \sum_{i=1}^N u_i^2 (x_i - \mu'_R)(x_i - \mu'_R)^T \quad (11)$$

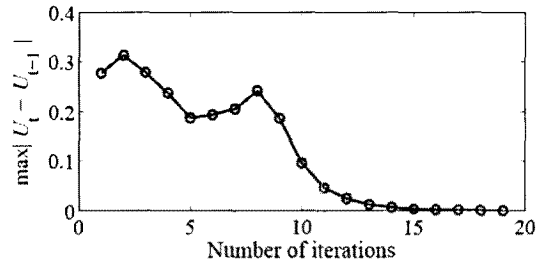
식 (11)와 (6)의 차이는 u_i 의 자승 유무이지만 중요한 의미를 갖는다. 부록에 주어진 증명을 통해 C' 의 고유 벡터는 식 (12)를 최적화함을 보일 수 있다.

$$J' = \sum_{i=1}^N u_i^2 \|(x_i - \mu'_R) - WW^T(x_i - \mu'_R)\|^2 \quad (12)$$

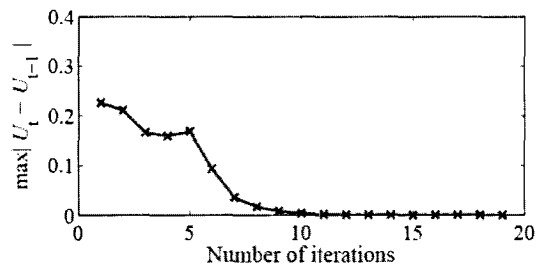
또한 식 (12)를 최적화하는 가중치 평균은 식 (4)가 아니라 식 (13)임을 알 수 있다.

$$\mu'_R = \frac{\sum_{i=1}^N u_i^2 x_i}{\sum_{i=1}^N u_i^2} \quad (13)$$

RF-PCA의 수렴성이 실험적으로 [5]에서 입증되었지만, 반복 과정에서 최대 소속도 차이가 영으로 수렴한다는 사실만이 입증되었을 뿐이다. 하지만 그림 2에 보여진 바와 같이 최대 소속도 차이는 RF-PCA와 RF-PCA2에서 항상 단조함수로 나타나는 것은 아니다. 따라서 RF-PCA에서 사용한 수렴 조건인 최대 소속도 차이를 사용하는 경우, 국부 최적치가 아닌 해를 구하는 경우가 발생할 수 있다. 표 1은 RF-PCA와 RF-PCA2에서 최적화하는 목적 함수를 비교하여 나타낸 것이다.



(a)



(b)

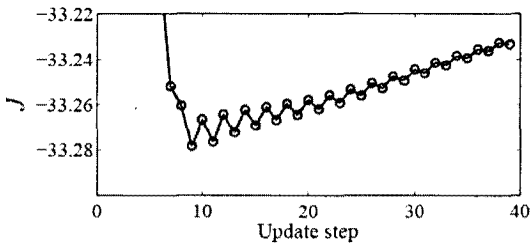
그림 2. 최대 소속도 차이의 예

(a) RF-PCA (b) RF-PCA2

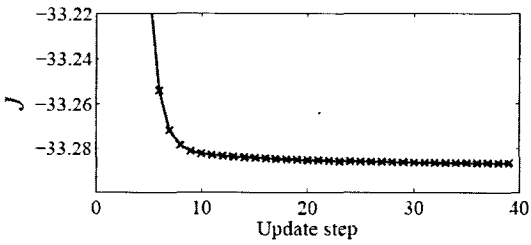
Fig. 2. An example of maximum membership difference (a) RF-PCA (b) RF-PCA2

표 1. RF-PCA와 RF-PCA2의 목적 함수 비교
Table 1. Objective functions of RF-PCA and RF-PCA2

RF-PCA	소속도 계산	$J = \sum_{i=1}^N u_i e(x_i) + \sigma^2 \sum_{i=1}^N (u_i \log u_i - u_i)$
	주성분 계산	$J'' = \sum_{i=1}^N u_i^2 e(x_i) + \sigma^2 \sum_{i=1}^N (u_i \log u_i - u_i)$
RF-PCA2	소속도 계산	$J = \sum_{i=1}^N u_i e(x_i) + \sigma^2 \sum_{i=1}^N (u_i \log u_i - u_i)$
	주성분 계산	



(a)



(b)

그림 3. 목적 함수 값의 변화 예
(a) RF-PCA (b) RF-PCA2

Fig. 3. An example of objective function value after each update (a) RF-PCA (b) RF-PCA2

그림 3은 그림 1에서 단계 5 및 6 이후 식 (1)의 목적 함수 값(J)을 나타낸 예이다. RF-PCA는 식 (1)과 (12)의 목적 함수를 번갈아 최적화함으로써, 그림 3에서와 같이 식 (1)의 목적 함수 값이 톱니 형태를 보일 수 있다. 이는 식 (12)의 목적 함수를 사용하는 경우도 마찬가지다. 하지만 RF-PCA2는 소속도와 고유치 모두 식 (1)의 목적 함수를 최적화함으로써 단조 감소 함수의 형태를 가지게 된다. 그림 2와 3을 위해 사용한 데이터는 그림 4에 나타나 있으며, 수렴 속도 및 구해진 주성분에 관해서는 3장에서 자세히 설명한다.

III. 실험 결과 및 고찰

RF-PCA2가 RF-PCA에 비해 나은 결과를 보인다는 것을 확인하기 위해 이 장에서는 RF-PCA와 RF-PCA2를 Matlab으로 구현하여 실험하였다. 그림 4는 실험에 사용한 데이터와 PCA 및 RF-PCA2를 사용하여 찾아낸 첫 번째 주성분을 표시한 것이다. 실험에 사용한 데이터는 2차원 가우시안 분포에서 생성한 타원형 형태의 데이터와, 역시 가우시안 분포에서 생성한 잡음으로 이루어져 있다. RF-PCA는 RF-PCA2와 유사한 주성분을 찾아냈기 때문에 그림 4에서는 표시하지 않았다. 그림 4에서 알 수 있듯이, PCA의 경우 잡음에 의해 원하는 주성분을 정확하게 찾아낼 수 없다. 하지만 RF-PCA2는 소속도를 통해 잡음에 작은 소속도 값을 할당함으로써 주성분을 정확하게 찾아낼 수 있다.

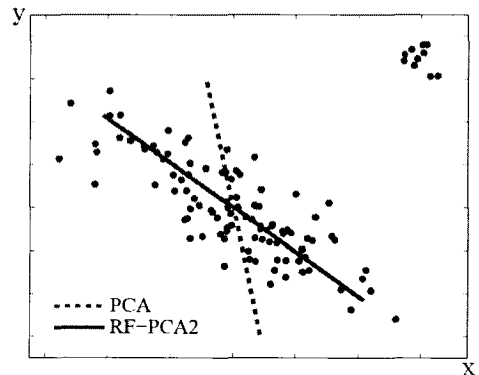
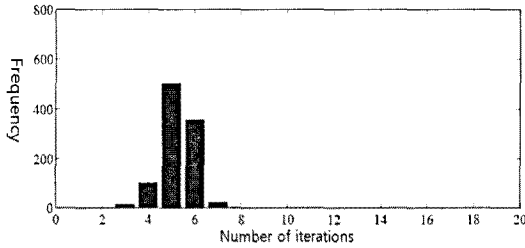


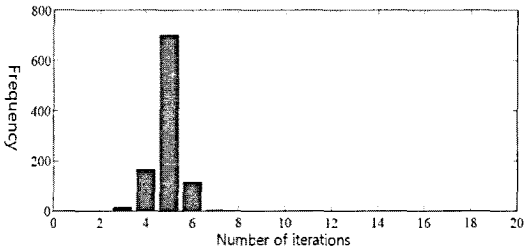
그림 4. PCA와 RF-PCA2를 이용하여 찾아낸 첫 번째 주성분

Fig. 4. First principal component found by PCA and RF-PCA2

첫 번째 실험에서는 RF-PCA와 RF-PCA2의 수렴 속도를 비교한다. 그림 5와 6은 RF-PCA와 RF-PCA2가 주어진 수렴 조건을 만족시키기 위해 필요한 반복 횟수를 나타낸 히스토그램이다. 실험은 1000회 반복하였으며, 2가지 수렴 조건을 사용하였다. 한 가지는 RF-PCA에서 사용한 조건으로 최대 소속도 차이를 수렴 조건으로 사용하였고, 다른 한 가지는 식 (1)의 목적 함수 차이를 사용하였다.



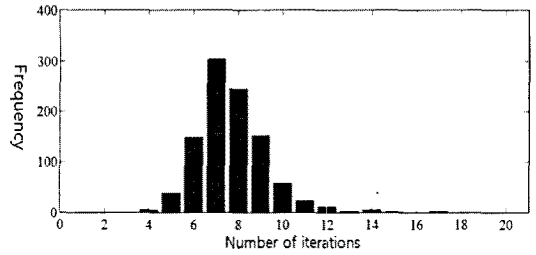
(a)



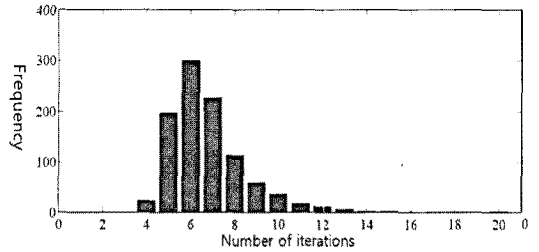
(b)

그림 5. 최대 소속도 차이 조건을 이용한 수렴 속도
(a) RF-PCA (b) RF-PCA2
Fig. 5. Convergence speed in using the maximum membership difference
(a) RF-PCA (b) RF-PCA2

표 2는 수렴 속도에 대한 실험 결과를 요약한 것으로, $\epsilon = 0.0001$, $t_{\max} = 20$ 으로 설정하였다. 이 실험에서 알 수 있듯이, RF-PCA2는 RF-PCA에 비해 두 가지 수렴 조건 모두에서 빠른 수렴 속도를 보인다. 한 가지 유의할 점은 수렴 조건 2에서 평균값이 크게 나타난 것은 조건 1과 조건 2의 단위가 달라서 이치 조건 1이 조건 2보다 좋은 조건임을 의미하지는 않는다는 점이다. 실제로 조건 1은 그림 2에 보여진 바와 같이 좋은 수렴 조건은 아니며, 이 논문에서는 RF-PCA와의 비교를 위해 사용하였다.



(a)



(b)

그림 6. 목적 함수 차이 조건을 이용한 수렴 속도
(a) RF-PCA (b) RKF-PCA
Fig. 6. Convergence speed in using the difference of objective function values
(a) RF-PCA (b) RKF-PCA

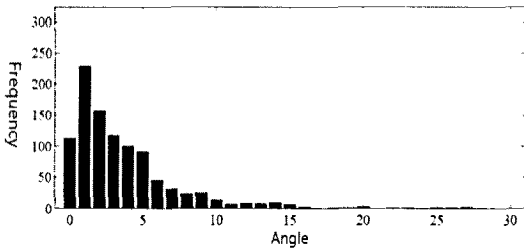
표 2. 평균 반복 회수
Table 2. Average number of iterations

조건	RF-PCA	RF-PCA2	
1	5.276	4.725	$\max U^{t-1} - U^t < \epsilon$
2	7.737	6.811	$J^{t-1} - J^t < \epsilon$

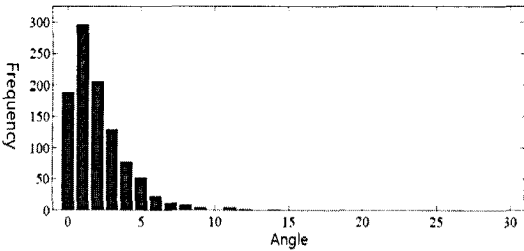
두 번째 실험은 RF-PCA와 RF-PCA2가 찾아낸 주성분의 질(quality)을 비교한다. 먼저 잡음이 없는 데이터에서 PCA를 이용하여 첫 번째 주성분 w_0 를 구하고, 잡음이 첨가된 데이터에서 RF-PCA와 RF-PCA2로 각각 첫 번째 주성분 w_1 과 w_2 를 구한다. 잡음이 첨가된 데이터에서 찾아낸 주성분이 잡음이 없는 경우의 주성분과 유사한 경우 그 성능이 우수한 것으로 생각할 수 있으므로, w_0 와 w_1 그리고 w_0 와 w_2 사이의 각도를 계산하여 비교하였다. 1000회 반복 실험한 각도의 평균은 RF-PCA의 경우 3.623°였고 RF-PCA2는 2.058°이었다. 즉, RF-PCA2가 RF-PCA에 비해 원하는 주성분을 더 정확히 찾아냄을 알

수 있다. 그림 7은 주성분 간 각도를 히스토그램으로 나타낸 것이다.

앞의 두 실험에서 알 수 있듯이, RF-PCA2는 RF-PCA의 단점인 공통의 목적 함수 부재를 해결함으로써 빠르고 정확한 주성분을 찾을 수 있음을 알 수 있다. 다른 실험 데이터에 대해서도 유사한 결과를 얻을 수 있었다.



(a)



(b)

그림 7. 두 주성분 사이의 각도

(a) RF-PCA (b) RF-PCA2

Fig. 7. Angles between two principal components

(a) RF-PCA (b) RF-PCA2

V. 결론

Robust fuzzy PCA(RF-PCA)는 퍼지 소속도를 이용하여 PCA의 잡음 민감성을 줄여주는 반복 알고리즘(iterative algorithm)이다. RF-PCA는 PCA에 비해 잡음 민감성이 줄어들지만 몇 가지 문제점이 있고, 수렴성이 그 중 하나이다. 이는 소속도와 주성분을 계산할 때 서로 다른 목적 함수를 사용하기 때문으로, 이 논문에서는 이러한 단점을 해결하기 위해 공통의 목적 함수를 최적화함

으로써 소속도와 주성분을 찾아낼 수 있는 RF-PCA2를 제안하였다. RF-PCA2는 RF-PCA에 비해 수렴 속도가 빠르며 구해진 해 역시 원하는 해에 보다 유사함을 실험 결과를 통해 알 수 있다.

RF-PCA2가 이전 방법에 비해 나은 결과를 보이지만, 개선의 여지는 여전히 남아있다. 그 중 한 가지는 커널을 이용한 비선형 알고리즘으로의 변환이다. PCA는 기본적으로 가우시안 분포만을 다룰 수 있다. 하지만 커널을 이용하는 경우 다양한 분포에 적용할 수 있으며, 현재 RF-PCA2의 비선형 커널 변형에 관해 연구 중에 있다. 또한 연산량을 줄이는 방법 역시 해결해야 할 문제 중 하나이다. RF-PCA2가 RF-PCA에 비해 연산량이 적은 연산량을 요구하지만 여전히 다른 PCA 변형에 비해 많은 연산을 요구하는 것은 사실이며, 이는 반복 알고리즘이라는 특징에 기인한다. 이를 해결하기 위해서는 매회 주성분을 다시 계산하지 않고 소속도의 차이만을 고려해줄 수 있는 incremental 알고리즘으로의 변형이 필요할 것으로 보이며, 이는 향후 연구 과제로 남아 있다.

부록

$\sum_{i=1}^N u_i x_i = 0$ 의 조건을 만족하는 D 차원 데이터 집합 $X = \{x_1, \dots, x_N\}$ 가 주어졌을 때, 잡음에 민감하지 않은 주성분을 찾아내기 위한 목적 함수는 식 (A-1)으로 나타낼 수 있다.

$$\arg \min_W J = \sum_{i=1}^N u_i \|x_i - WW^T x_i\|^2 \quad (A-1)$$

이 때 N 은 데이터 포인트의 개수를, u_i 는 x_i 의 소속도를, W 는 직교 단위 벡터로 이루어지는 행렬을 나타낸다. $W = [w_1, \dots, w_D]$ 가 식 (A-2)를 만족시키는 D 차원의 직교 단위 벡터라고 가정하자.

$$w_i^T w_j = \begin{cases} 1, & \text{if } i=j \\ 0, & \text{otherwise} \end{cases} \quad (A-2)$$

W 를 이용하여 데이터 포인트 x_i 는 직교 단위 벡터의 가중치 합으로 식 (A-3)과 같이 표현될 수 있다.

$$x_i = \sum_{d=1}^D (x_i^T w_d) w_d \quad (\text{A-3})$$

또한 k 개의 직교 단위 벡터로 재건된(reconstructed) 데이터 포인트는 식 (A-4)로 표현된다.

$$\tilde{x}_i = \sum_{d=1}^k a_{id} w_d + \sum_{d=k+1}^D b_d w_d \quad (\text{A-4})$$

식 (A-3)과 (A-4)를 이용하여 목적 함수는 식 (A-5)로 변환될 수 있다.

$$\begin{aligned} J &= \sum_{i=1}^N u_i \|x_i - WW^T x_i\|^2 \quad (\text{A-5}) \\ &= \sum_{i=1}^N u_i \left(x_i - \sum_{d=1}^k a_{id} w_d - \sum_{d=k+1}^D b_d w_d \right)^2 \\ &= \sum_{i=1}^N u_i \left(\begin{array}{l} x_i^T x_i - 2x_i^T \sum_{d=1}^k a_{id} w_d + \sum_{d=1}^k a_{id}^2 \\ - 2x_i^T \sum_{d=k+1}^D b_d w_d + \sum_{d=k+1}^D b_d^2 \end{array} \right) \end{aligned}$$

$u_i \neq 0$ 인 경우, 식 (A-5)를 a_{id} 에 따라 편미분하면 식 (A-6)을 얻을 수 있다.

$$\frac{\partial J}{\partial a_{id}} = u_i (-2x_i^T w_d + 2a_{id}) = 0 \quad (\text{A-6a})$$

$$a_{id} = x_i^T w_d \quad (\text{A-6b})$$

따라서 a_{id} 의 최적값은 x_i 를 벡터 w_d 로 투사한 값을 알 수 있다. 유사하게 식 (A-5)를 b_d 에 따라 편미분하면 식 (A-7)을 얻을 수 있다.

$$\frac{\partial J}{\partial b_d} = \sum_{i=1}^N u_i (-2x_i^T w_d + 2b_d) = 0 \quad (\text{A-7a})$$

$$b_d = \left(\frac{\sum_{i=1}^N u_i x_i}{\sum_{i=1}^N u_i} \right)^T w_d = 0 \quad (\text{A-7b})$$

따라서 b_d 의 최적값은 0임을 알 수 있다. 식 (A-6b)와 (A-7b)를 이용하면, 데이터 포인트와 재건된 포인트의 차이는 식 (A-8)과 같이 나타내어진다.

$$\begin{aligned} x_i - \tilde{x}_i &= \sum_{d=1}^D (x_i^T w_d) w_d - \sum_{d=1}^k (x_i^T w_d) w_d \quad (\text{A-8}) \\ &= \sum_{d=k+1}^D (x_i^T w_d) w_d \end{aligned}$$

식 (A-8)을 이용하여 식 (A-1)의 목적 함수는 식 (A-9)로 나타낼 수 있다.

$$\begin{aligned} J &= \sum_{i=1}^N u_i \|x_i - WW^T x_i\|^2 \quad (\text{A-9}) \\ &= \sum_{i=1}^N u_i \left(\sum_{d=k+1}^D (x_i^T w_d) w_d \right)^2 \\ &= \sum_{i=1}^N u_i \sum_{d=k+1}^D (x_i^T w_d)^2 \\ &= \sum_{d=k+1}^D \sum_{i=1}^N u_i w_d^T x_i x_i^T w_d \\ &= \sum_{d=k+1}^D w_d^T \left(\sum_{i=1}^N u_i x_i x_i^T \right) w_d \\ &= \sum_{d=k+1}^D w_d^T S w_d \end{aligned}$$

이 때 S 는 가중치 분포 행렬(weighted scatter matrix)이라 한다. 벡터 w_d 의 크기는 1이 되어야 한다는 조건을 이용하면 라그랑지 방정식은 식 (A-10)으로 나타낼 수 있다.

$$L = \sum_{d=k+1}^D w_d^T S w_d + \sum_{d=k+1}^D \lambda_d (1 - w_d^T w_d) \quad (\text{A-10})$$

이 때 λ_d 는 라그랑지 상수를 나타낸다. 식 (A-10)을 w_d 에 의해 편미분하면, 식 (A-11)을 얻을 수 있다.

$$\frac{\partial L}{\partial w_d} = w_d^T (S^T + S) - 2\lambda_d w_d = 0 \quad (\text{A-11a})$$

$$S w_d = \lambda_d w_d \quad (\text{A-11b})$$

따라서 w_d 의 최적값은 S 의 고유 벡터임을 알 수 있다. 식 (A-9)와 (A-11b)로부터 식 (A-1)의 목적함수는 식 (A-12)로 변형될 수 있다.

$$J = \sum_{d=k+1}^D \lambda_d \quad (A-12)$$

식 (12)로부터 J 의 최소값은 $D-k$ 개의 최소 고유치를 합한 값이며, 최적의 k 개 작고 단위 벡터는 S 의 고유 벡터 중 k 개의 최대 고유치에 해당하는 벡터들임을 알 수 있다.

참고문헌

[1] I. T. Jolliffe, *Principal Component Analysis*, 2nd Edition, Springer, 2002.

[2] P. Rousseeuw, "Multivariate estimation with high breakdown point," *Mathematical Statistics and Applications B*, pp. 283 - 297, 1985.

[3] C.-D. Lu, T.-Y. Zhang, X.-Z. Du, and C.-P. Li, "A robust kernel PCA algorithm," *Proceedings of the 3rd International Conference on Machine Learning and Cybernetics*, pp. 3084 - 3087, 2004.

[4] C. Lu, T. Zhang, R. Zhang, and C. Zhang, "Adaptive robust kernel PCA algorithm," *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing*, pp. VI 621 - 624, 2003.

[5] G. Heo, P. Gader, and H. Frigui, "RKF-PCA: Robust kernel fuzzy PCA," *Neural Networks* Vol. 22, No. 5-6, pp. 642 - 650, 2009.

[6] T.-N. Yang and S.-D. Wang, "Fuzzy auto-associative neural networks for principal component extraction of noisy data," *IEEE Transaction on Neural Networks*, Vol. 11, No. 3, pp. 808 - 810, 2000

[7] T. R. Cundari, C. Sarbu, and H. F. Pop, "Robust fuzzy principal component analysis (FPCA). A comparative study concerning interaction of carbonhydrogen bonds with molybdenum-oxo bonds," *Journal of Chemical Information and Computer Sciences*, Vol. 42, No. 6, pp. 1363 - 1369, 2002.

[8] J. C. Bezdek and R. J. Hathaway, "Convergence of alternating optimization," *Neural, Parallel and Scientific Computations*, Vol. 11, No. 4, pp. 351 - 368, 2003.

[9] G. Heo and P. Gader, "Fuzzy SVM for noisy data: A robust membership calculation method," *Proceedings of the 2009 IEEE International Conference on Fuzzy Systems*, pp. 431 - 436, 2009.

[10] H. Ichihashi, K. Honda, and N. Tani, "Gaussian mixture PDF approximation and fuzzy c-means clustering with entropy regularization," *Proceedings of the 4th Asian Fuzzy Systems Symposium*, pp. 217 - 221, 2000.

[11] P. J. Huber, *Robust Statistics*, Wiley- Interscience, 1981.

저자소개

허경용(Gyeongyong Heo)



- 1996년 8월 : 연세대학교 본대학원 전자공학과 (공학석사)
- 2009년 12월: Department of Computer and Information Science

and Engineering, University of Florida (공학박사)

※ 관심분야: Machine Learning, Pattern Recognition, Image Processing

우영운(Young Woon Woo)



- 1989년 2월 : 연세대학교 전자공학과(공학사)
- 1991년 8월 : 연세대학교 본대학원 전자공학과(공학석사)

- 1997년 8월 : 연세대학교 본대학원 전자공학과 (공학박사)
- 1997년 9월 ~ 현재: 동의대학교 멀티미디어공학과 교수

• 2010년 ~ 현재: 한국해양정보통신학회 국제학술 분과위원장

※ 관심분야: 지능시스템, 패턴인식, 퍼지이론, 의료정보



김성훈(Seong Hoon Kim)

- 1996년 2월: 연세대학교 본대학원 전자공학과 (공학박사)
 - 1996년 3월 ~ 2006년 2월: 영동대학교 컴퓨터공학과 부교수
 - 2006년 3월 ~ 현재: 경북대학교 컴퓨터정보학부 조교수
- ※ 관심분야: 인공지능, 패턴인식, 지능형콘텐츠