

# PCA 혼합 모형을 위한 효율적인 구조 선택 방법

김현철<sup>0</sup> 김대진 방승양  
포항공과대학교 컴퓨터공학과  
(grass, dkim, sybang}@postech.ac.kr

## An Efficient Model Selection Method for a PCA Mixture Model

Hyun-Chul Kim<sup>0</sup> Daijin Kim Sung-Yang Bang  
Dept. of Computer Science and Engineering, POSTECH

### 요 약

PCA는 다변수 데이터 해석법 중 가장 널리 알려진 방법 중 하나로 많은 응용을 가지고 있다. 그런데, PCA는 선형 모델이어서 비선형 구조를 분석하는데 효과적이지 않다. 이를 극복하기 위해서 PCA의 조합을 이용하는 PCA 혼합 모형이 제안되었다. PCA 혼합 모형의 핵심은 구조 선택, 즉 mixture 요소의 수와 PCA 기저의 수의 결정인데 그의 체계적인 결정 방법이 필요하다. 본 논문에서는 단순화된 PCA 혼합 모형과 이를 위한 효율적인 구조 선택 방법을 제안한다. 각각의 mixture 요소 수에 대해서 모든 PCA 기저를 갖도록 한 상태에서 PCA 혼합 모형의 파라미터를 EM 알고리즘을 써서 결정한다. 최적의 mixture 요소의 수는 오류를 최소로 하는 것으로 결정한다. PCA 기저의 수는 PCA의 정렬성 특성을 이용해서 중요도가 적은 기저부터 하나씩 잘라 내며 오류가 최소로 하는 것으로 결정한다. 제안된 방법은 특히 다차원 데이터의 경우에 EM 학습의 횟수를 많이 줄인다. 인공 데이터에 대한 실험은 제안된 방법이 적절한 모델 구조를 결정한다는 것을 보여준다. 또, 눈 감지에 대한 실험은 제안된 방법이 실용적으로도 유용하다는 것을 보여준다.

### 1. 서론

PCA는 다변수 해석 방법으로 잘 알려진 기법이다 [1]. PCA의 특성은 데이터를 가능한 한 잘 보존하면서 데이터의 차원을 줄이는 것이다. 이러한 특성으로 인해 데이터 압축, 영상해석, 시각화, 패턴인식, 회귀, 시계열예측 등에 응용되었다. 그런데, PCA는 선형성의 특성으로 비선형 데이터 분포를 모델링하는 데는 한계가 있다.

이를 극복하기 위해서 여러 개의 PCA를 결합한 PCA 혼합 모형을 사용할 것을 제안한 여러 연구가 시도되었다. Hinton[2] 등은 복귀예러를 이용하는 PCA 혼합 모형을 제안하였다. Tipping과 Bishop[3]은 확률적 PCA의 혼합 모형을 제안하였다. 여기서는 Tipping과 Bishop의 모델을 단순화시킨 PCA 혼합 모형을 제안하고, 효율적인 모델 선택을 위한 준최적화 선택방법을 제안한다.

본 논문은 다음과 같이 구성된다. 2장에서는 PCA 혼합 모형에 대해서 설명하고 3장에서는 학습 및 모델 선택 방법을 제시한다. 4장에서는 인공데이터와 눈 감지에 대한 실험결과를 설명하며, 5장에서는 결론이 뒤따른다.

### 2. PCA 혼합 모형

PCA 혼합 모형을 이용해서 주어진 데이터를 어떻게 표현하는지 살펴보도록 하겠다. 주어진 하나의  $N$ 차원 데이터

$x$ 를 다음과 같이 표현한다.

$$x = T_k s_k + m_k$$

여기서  $T_k$ 는 각 열벡터가 PCA 기저인  $N \times N$  행렬이다.  $s_k$ 는 데이터에 대한 PCA기저 계수로  $N$ 차원 벡터이다.  $m_k$ 는 bias로  $N$ 차원 벡터이다.  $k$ 는  $k$ 번째 mixture 요소를 나타낸다.

그림1은 임의의 분포를 갖는 2차원 데이터  $x$ 를 나타낸 것이다. 이 그림에서 혼합 요소 수는 2개이다. 점들의 군집 내에 표시된 선분은  $T_k$ 의 열벡터의 방향 즉, PCA 기저의 주축 방향을 의미한다. 두 선분의 교차점들은  $m_k$ 를 나타낸다.

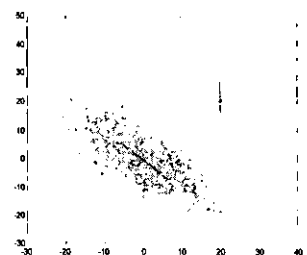


그림 1 PCA 혼합 모형의 설명

각 mixture 요소에 해당하는 군집을 하나의 class로 보고 이를  $C_k$ 라고 하자. 그러면 데이터  $x$ 가 각 class에 속할 확률은 아래와 같이 표현될 수 있다. 여기서 각 class의 분포는 Gaussian 분포함수라고 가정하였다.

$$p(\mathbf{x} | \theta_k, C_k) = \prod_{l=1}^N \left[ \frac{1}{(2\pi)^{1/2} \lambda_{kl}^{1/4}} \exp \left\{ -\frac{((s_k)_l)^2}{2\lambda_{kl}} \right\} \right]$$

$$(S_k = \mathbf{T}_k^T (\mathbf{x} - \mathbf{m}_k))$$

$$p(C_k | \mathbf{x}, \theta_{1:N}) = \frac{p(\mathbf{x} | \theta_k, C_k) p(C_k)}{\sum_{k=1}^N p(\mathbf{x} | \theta_k, C_k) p(C_k)}$$

$$(\theta_k = \{\mathbf{T}_k, \mathbf{m}_k, \{\lambda_{kl} | l=1,2,\dots,N\}\})$$

### 3. 파라미터 학습과 구조 선택

PCA 혼합 모형은 두 가지를 결정해야 한다. 하나는 모델의 파라미터이다. 그런데 이 파라미터를 얻기 위해서는 확률 분포가 필요하다. 이러한 경우에 EM 알고리즘을 이용해서 모델의 파라미터와 확률 분포를 추정할 수 있다는 것이 알려져 있다. PCA 혼합 모형을 위한 EM 학습 알고리즘은 아래와 같다. E-step과 M-step을 수렴할 때까지 반복한다. 이 EM 학습 알고리즘은 최대 유사도 원리에 의해 유도될 수 있다.

#### ● E-step

$$m_k = \frac{\sum_{i=1}^I \mathbf{x}_i p(C_k | \mathbf{x}_i, \theta_{1,K})}{\sum_{i=1}^I p(C_k | \mathbf{x}_i, \theta_{1,K})}$$

-PCA

$$Cov(k) = E(p(\mathbf{x} | C_k, \theta_k)(\mathbf{x} - \mathbf{m}_k)^T (\mathbf{x} - \mathbf{m}_k))$$

$$\lambda_{kl} = \text{eigenvalue}(Cov(k))_l$$

$$(\lambda_{k1}, \lambda_{k2}, \dots, \lambda_{kN} : \text{decreasing order})$$

$$\mathbf{T}_k = [\mathbf{v}_{k1} \ \mathbf{v}_{k2} \ \dots \ \mathbf{v}_{kN}]$$

$$(\mathbf{v}_{kl} = \text{normalized\_eigenvector}(Cov(k))_{\lambda_{kl}})$$

#### ● M-step

$$p(\mathbf{x} | \theta_k, C_k) = \prod_{l=1}^N \left[ \frac{1}{(2\pi)^{1/2} \lambda_{kl}^{1/4}} \exp \left\{ -\frac{((s_k)_l)^2}{2\lambda_{kl}} \right\} \right]$$

$$(S_k = \mathbf{T}_k^T (\mathbf{x} - \mathbf{m}_k))$$

$$p(C_k | \mathbf{x}, \theta_{1:N}) = \frac{p(\mathbf{x} | \theta_k, C_k) p(C_k)}{\sum_{k=1}^N p(\mathbf{x} | \theta_k, C_k) p(C_k)}$$

$$(\theta_k = \{\mathbf{T}_k, \mathbf{m}_k, \{\lambda_{kl} | l=1,2,\dots,N\}\})$$

학습과 관련해서 두 번째 결정해야 하는 것은 mixture 요소의 수와 PCA기저들의 수이다. 너무 많으면 데이터 분포를 overfitting할 수 있고 너무 적으면 데이터 분포를 잘 추정할 수 없다. 문제를 간단히 하기 위해, 모든 부류는 같은 수의 mixture 요소를 갖는다고 가정하고 모든 mixture 요소는 같은 수의 PCA 기저를 사용한다고 가정한다. 데이터에 대한 검증집합(validation set)에 대해 복귀오차가 최소인 것으로 mixture 요소 수와 PCA기저 수

를 정하는 방법을 사용한다.

PCA 기저는 중요도(significance)순서로 배열되는 특성을 가지고 있어서 중요도가 적은 것부터 잘라내면서 각 PCA기저 수에 대한 검증집합에 대해 복귀오차를 계산할 수 있다. 구체적인 알고리즘은 아래 그림2와 같다.

```

for k=1 to  $K_{max}$ 
  for  $C=C_1$  to  $C_{max}$ 
    Compute the complete transform matrix
     $W_{C,k}(1:n)$  using the training data set  $\mathcal{T}_R$ .
  end
  for m=n to 1 by -1
    Compute classification error(k,m)
    using the validation data set  $\mathcal{T}_V$ .
  end
end
end
( $K_C, m_C$ ) =  $\underset{k,m}{\text{argmin}}$ (classification error(k,m)).
    
```

그림 2 최적화 모델 선택 방법

이런 구조 선택 방법은 기존의 방법보다 시간이 단축된다. 즉, 우선 PCA기저 수를 데이터의 차원과 같도록 해서 EM 학습을 한 후에 PCA기저의 수를 결정한다. 그래서 EM 학습 전에 PCA기저의 수를 결정하는 방법에 비해 EM 학습의 수를 줄일 수 있는 장점이 있다.

### 4. 실험 결과 및 분석

#### 4.1. 인공데이터에 대한 실험

제안된 모델과 구조 선택 방법이 잘 동작하는지 확인해보기 위해서 많은 실험을 했다. 여기서는 두 가지 실험 결과를 제시한다.

첫번째 예로, 각각의 군집이 가우스 분포함수에 의해 생성되고 각 부류가 5개의 군집을 갖는 2부류의 데이터에 대해서 실험하였다. 각 군집에 대해 1,000개의 데이터가 생성하여서 전체 데이터 수는 10,000개이다. 전체 데이터를 4,423개의 훈련집합, 2,253개의 검증집합, 3,319개의 테스트집합으로 나누었다.

각각의 부류를 PCA 혼합 모형에 의해서 모델링하였다. 앞에서 제시된 파라미터 학습과 구조 선택방법에 의해 파라미터를 학습했고 구조를 선택했다. 구조 선택을 위해서 여러 수의 mixture 요소에 대해서 파라미터 학습을 해 본 후 검증집합에 대한 오류를 계산하였다. 그 결과는 아래 표1과 같다.

Number of mixture components	1	2	3	4
Validation set error(%)	31.0	13.2	4.44	3.68
Number of mixture components	5	6	7	8
Validation set error(%)	1.78	1.95	2.22	2.09

표 1 첫 번째 인공데이터에 대한 검증집합 오류

보는 바와 같이 mixture 요소의 수가 5일 때 가장 좋은 성능을 보였다. 이는 각각의 부류가 5개의 군집

으로 되어 있다는 사실과 일치한다. 이로부터 제안된 구조 선택방법을 이용해서 최적의 mixture 요소 수를 얻을 수 있음을 알 수 있다. 한편, mixture 요소의 수가 5일 때 테스트 오류는 1.42%이다.

두 번째 예로, 한 부류가 3개의 PCA기저를 갖는 2개의 군집으로 구성된 2부류의 데이터에 대해서 실험하였다. 각 군집 마다 1000개의 데이터가 생성되어서 전체 데이터 수는 4000개가 되었다. 이 데이터를 1796개의 훈련 집합, 926개의 검증집합, 1278개의 테스트 집합으로 나누었다.

각각의 부류를 PCA 혼합 모형에 의해서 모델링하였다. 앞에서 제시된 파라미터 학습과 구조 선택방법에 의해 파라미터를 학습했고 구조를 선택했다. 구조 선택을 위해서 2개의 mixture 요소에 대해 파라미터 학습을 해 본 후 여러 수의 PCA기저 수에 대해서 검증 집합에 대한 오류를 계산하였다. 그 결과는 아래 표2과 같다.

Number of PCA bases	1	2	3	4	5
Validation set error(%)	2.81	3.35	0	2.70	22.03

표 2 두 번째 인공데이터에 대한 검증집합 오류

보는 바와 같이 PCA기저의 수가 3일 때 가장 좋은 성능을 보였다. 이는 각 부류의 각 군집이 3개의 PCA기저로 되어 있다는 사실과 일치한다. 이를 볼 때 제안된 구조 선택방법으로 최적의 PCA기저 수를 얻을 수 있음을 알 수 있다. 한편, PCA기저의 수가 3일 때 테스트 오류는 0%이다.

#### 4.2. 눈 감지 실험

눈 감지에는 여러 방법이 제안되었는데, 그 중에 고유공간 방법(eigenspace method)이 있다[4]. PCA의 효과를 이용해서 얼굴영상의 여러 패치에 대해 복구사 근사화오차(reconstruction error)가 가장 작은 패치의 지점을 찾음으로써 눈을 감지하는 방법이다. 제안된 방법을 이용하여 눈 감지에 적용해 보았다. 즉, PCA를 이용하는 고유공간방법을 PCA 혼합 모형을 이용하여 확장해 보았다. 데이터 베이스는 CMU에서 제공하는 얼굴 영상 집합에 CMU의 얼굴감지 방법을 적용해 얻어진 얼굴이미지를 사용하였다 [5]. 조명변화를 극복해보기 위한 방법으로 histogram-equalization을 수행하였다.

얼굴 영상의 크기는 50x50이고 다양한 포즈나 표정, 안경착용, 조명 변화 등을 포함한다. PCA 혼합 모형을 학습을 위해 190개의 영상을 사용하였다. 또, 테스트를 위해 100개의 영상을 사용하였다.

여기서 훈련을 위해서 눈을 잘라낸 후 8x8의 크기로 크기를 조정했다. 이것들을 이용하여 각 눈에 대한 PCA 혼합 모형의 파라미터를 EM학습 알고리즘에 의해 추정하였다. 검증집합으로는 훈련집합의 첫 10개를 사용했다. 검증 결과 mixture 요소가 2개이고, PCA기저는 25개일 때 가장 좋은 성능을 보였다. PCA만을 이용할 때는 PCA기저 수가 18일 때 좋은 성능을 보였다.

눈 감지는 다음과 같이 이루어진다. 눈 후보 영역(왼쪽 눈의 경우 왼쪽 위 1/4영역, 오른쪽 눈의 경우

오른쪽 위 1/4영역)에서 위치와 크기에 따라 패치를 얻는다. 각 mixture 요소별로 복구 오차를 계산한다. 얻어진 패치를 8x8로 조정된 후 이를 모든 mixture 요소에 대해서, 또 모든 크기와 모든 위치에 대해서 계산한 후, 최소의 복구에러를 주는 것을 찾는다. 이에 해당하는 위치와 크기에 해당하는 패치를 눈으로 감지한다.

실험 결과는 표 3과 같다. 눈의 중심과 패치의 중심의 차가 4-5 픽셀이하인 경우를 성공한 경우로 간주했다. 그림 3은 테스트 집합에 대해 눈을 감지한 예를 나타낸 것이다. 눈 감지 성능을 보면 PCA를 사용하는 고유공간방법보다 제안된 모델을 이용하는 것이 더 좋은 성능을 보임을 알 수 있다.

Method	Left eye (%)	Right eye (%)
PCA	69	72
PCA mixture	80	76

표 3 눈 감지율 (eye detection rate)

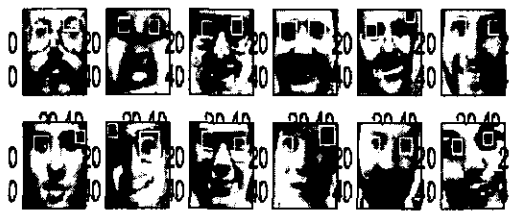


그림 3 눈 감지 결과 예

#### 6. 결론

이 논문은 단순화된 PCA 혼합 모형과 이를 위한 구조 선택 방법을 제시한다. 인공 데이터에 대한 실험은 제안된 방법이 실제 최적의 mixture 요소 수와 최적의 PCA기저 수를 결정함을 보여준다. 눈 감지에 대한 실험에서 PCA만을 사용하는 고유공간방법보다 더 좋은 성능을 보여주었다. 이를 볼 때 제안된 방법은 실용적으로도 가치가 있음을 알 수 있다.

#### 7. 참고 문헌

- [1] Jolliffe, I. T., Principal component analysis, New York : Springer-Verlag
- [2] Hinton, G.E., Dayan, P., Revow, M. "Modeling the manifolds of images of handwritten digits", IEEE Transactions on Neural Networks, 8(1), pp. 65-74, 1997
- [3] Tipping, M. E., Bishop, C. M., "Mixtures of Probabilistic Principal Component Analyzers", Neural Computation, 11, pp. 443-482, 1999
- [4] Pentland, A., Moghaddam, B., Starner, T., "View-based and modular eigenspaces for face recognition", Proceedings CVPR '94, pp. 84-91, 1994
- [5] Schneiderman, H., Kanade, T., "Probabilistic Modeling of Local Appearance and Spatial Relationships for Object Recognition." IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 45-51, 1998