

# Sparse NMF에 의한 클러스터링

## Clustering Effects in Sparse NMF(Non-negative Matrix Factorization)

오상훈

목원대학교

Sang-Hoon Oh

Mokwon Univ.

### 요약

입력에서 특징을 추출하는 유용한 방법으로 NMF(Non-negative Matrix Factorization)이 제안되었다. NMF를 적용하면 고차원의 데이터가 저차원의 특징에 기반한 형태로 변형이 된다. 이 경우 클러스터링 효과도 같이 나타나는데, 최근에 Sparse NMF가 이러한 효과를 더 잘 보인다고 알려졌다. 이 논문에서는 숫자 영상 신호에 대하여 NMF와 Sparse NMF를 적용시켜 이러한 클러스터링 효과를 비교하여 보았다.

### Abstract

NMF(Non-negative Matrix Factorization) has been proposed as an useful algorithm for feature extraction. Using NMF, we can extract low-dimensional feature vectors. Also, we can find clustering effects in the NMF algorithm. Also, it is reported that the sparse NMF algorithm shows better clustering effects. This paper compares the two approaches in the viewpoint of clustering effects.

## I. 서론

고차원의 데이터를 저차원으로 줄이는 문제는 많은 응용 분야를 지니기 때문에 중요한 연구분야로 다루어져 왔다. 이러한 차원 축소에는 크게 두가지(특징추출(feature extraction)과 특징 선택(feature selection))의 접근 방식이 있다. 특징추출은 고차원의 데이터를 저차원의 데이터로 변환시키는 것이고, 특징선택은 변환 없이 고차원 데이터의 부분집합을 선택하는 것이다. 특징추출을 위하여 PCA(Principal Component Analysis), LDA(Linear Discriminant Analysis), NMF(Non-negative Matrix Factorization)[1] 등이 많이 사용된다. 특히, 입력 데이터가 양의 값만을 가지는 영상을 대상으로 하는 경우에는 NMF가 유용한 방법이다.

NMF는 고차원의 데이터 행렬을 저차원의 basis 행렬과 coefficient 행렬의 곱으로 표시한다. NMF의 특징 중 하나는 sparse basis vectors를 얻을 수 있다는 것이고, 이를 기반으로 parts-based basis vectors를 얻을 수 있다. Basis와 Coefficient 행렬의 sparseness를 조절한다면 NMF의 더 유용한 결과를 얻을 수 있을 것이라는 기대로 많은 연구가 진행되었으며, 특히 Sparse NMF[2]가 주목할 만하다. 또한, ICA(Independent Component Analysis) 역시 입력의 특징을 추출해내는 기능이 있으므로[3], ICA에 의한 특징과 NMF에 의한 특징을 비교하는 것도 특징추출의 관점에서 유용하다.

이 논문은 NMF와 Sparse NMF를 필기체 숫자 영상을 대상으로 시뮬레이션하여 Sparseness 때문에 나타나는 clustering

효과를 조사해보고자 한다. 또한, NMF와 Sparse NMF에 의해 추출된 특징과 ICA에 의해 추출된 특징을 비교해보고자 한다.

## II. NMF, Sparse NMF, ICA 알고리즘 소개

### 1. NMF(Non-negative Matrix Factorization)

부분의 특징에 근거한 인식을 두뇌가 수행한다는 것에 착안하여, 이를 알고리즘으로 구현한 것이 NMF이다. 특히, 특징들은 음이 아닌 영 혹은 양의 값만 지닌다는 것을 중시한 것이다.

$n \times m$  영상 행렬  $V$ 가 주어졌다고 하자. 각 행은  $n$  개의 양의 값을 지닌 pixel을 나타내며,  $m$  은 영상의 개수이다. NMF는  $V$ 를

$$V = WH \quad (1)$$

와 같이 변환시킨다. 즉,

$$\min f(W, H) = \frac{1}{2} \| V - WH \|_F^2 \quad (2)$$

인  $W$ 와  $H$ 를 구한다. 여기서,  $W$ 는  $n \times r$  이고  $H$ 는  $r \times m$  이다. 이때, rank  $r$  은 일반적으로  $(n+m)r < nm$  이다. 식 (1)은 일종의 축약된 형태의 데이터 표현이라고 볼 수 있다. 이 축약된 형태의  $W$ 와  $H$ 는

$$F = \sum_{i=1}^n \sum_{\mu=1}^m [V_{i\mu} \log(WH)_{i\mu} - (WH)_{i\mu}] \quad (3)$$

를 최대화 시키는 방향으로 반복적으로 변경이 된다[1].

## 2. Sparse NMF

NMF의 특징 중 하나는 sparse를 부여하여 parts-based basis vectors를 얻게 해주는 것이다. 이를 위하여 NMF알고리즘에 sparseness를 부가하기 위하여 여러 가지 알고리즘이 제안되었다. 그 중 하나인 H에 sparseness를 부여한 알고리즘을 이 논문에서 사용하고자 한다. 즉, H에 sparseness를 부여하여 클러스터링 효과를 보고자 하기 때문이다. 이를 위한 알고리즘은

$$\min f(\mathbf{W}, \mathbf{H}) = \frac{1}{2} \|\mathbf{V} - \mathbf{WH}\|_F^2 + \eta \|\mathbf{W}\|_F^2 + \beta \sum_{j=1}^m \|\mathbf{H}(:, j)\|_1^2 \quad (4)$$

를 최소화 시키는 것이다[2].

## 3. ICA(Independent Component Analysis)

독립 성분 분석(ICA: Independent Component Analysis)은 선형적으로 신호를 해석하는 하나의 방법으로서, 먼저 source signal vector는

$$\mathbf{s} = [s_1, s_2, \dots, s_N]^T \quad (5)$$

와 같이 주어진다고 하자. 여기서  $s_i$ 는 서로 독립이라고 하며, 그 분포는 non-Gaussian 이거나 아니면 오직 한 source만 Gaussian 이라고 한다. 이 가정 하에서 우리가 측정하는 신호는

$$\mathbf{x} = \mathbf{A} \mathbf{s} \quad (6)$$

와 같이 주어진다고 하자. 이때  $\mathbf{A}$ 는  $N \times N$  mixing matrix라고 한다.

위와 같은 가정 하에,  $\mathbf{s}$ 에 관한 정보가  $s_i$ 가 서로 독립이라는 것 이외에는 전혀 없는 상황에서, 측정된 신호  $\mathbf{x}$ 로부터 다시 source signal vector  $\mathbf{s}$ 를 복원해 내고자 한다. 이러한 목적으로

$$\mathbf{u} = \mathbf{W} \mathbf{x} \quad (7)$$

와 같이 변환 한 후,  $y_i = g(u_i)$  형태로 비선형 변환을 거친다. 이때,  $\mathbf{W}$ 는  $N \times N$  unmixing matrix라고 하며,  $g(\cdot)$ 는

source signal의 cumulative density function이 되도록 한다. 그러면, 이 network의 출력이 지닌 joint entropy가 최대가 되도록 학습을 한다면  $\mathbf{u}$ 는 독립이 되도록 학습된 것이며, source signal  $\mathbf{s}$ 를 찾아낸 것이다. 즉,

$$\frac{\partial H(\mathbf{y})}{\partial \mathbf{W}} = [(\mathbf{W}^T)^{-1} - \phi(\mathbf{u}) \mathbf{x}^T] \quad (8)$$

에 의해 계산된 결과에서 gradient가 증가하는 방향으로  $\mathbf{W}$ 를 변경시킨다. 이를 더 효율적으로 하기 위해 제안된 "natural gradient"를 적용하면

$$\Delta \mathbf{W} \propto \frac{\partial H(\mathbf{y})}{\partial \mathbf{W}} \mathbf{W}^T \mathbf{W} = [\mathbf{I} - \phi(\mathbf{u}) \mathbf{u}^T] \mathbf{W} \quad (9)$$

와 같이 된다. 여기서,

$$\phi(\mathbf{u}) = - \frac{\frac{\partial p(\mathbf{u})}{\partial \mathbf{u}}}{p(\mathbf{u})} \quad (10)$$

는 "score function"이라고 하며,  $p(\cdot)$ 는 source signal의 p.d.f.(probability density function)이다.

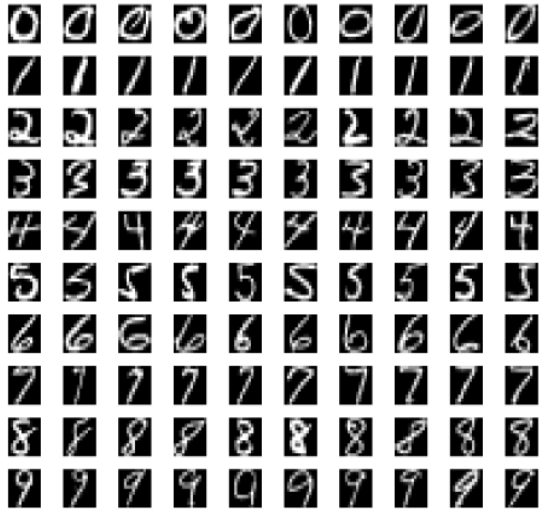
위와 같이 유도된 ICA 알고리즘은 출력의 entropy를 최대로 한다는 뜻에서 Information Maximization을 줄여 "InfoMax" 방법이라고 부른다. 이와 달리, 출력들 간의 mutual information을 최소화 시키도록 학습하는 "Minimizing Mutual Information" 알고리즘도 있다. 이 두 방법은 한 목적을 다른 관점에서 접근한 것으로 그 결과는 식 (9)와 같은 형태로 나타난다[3].

## III. 시뮬레이션

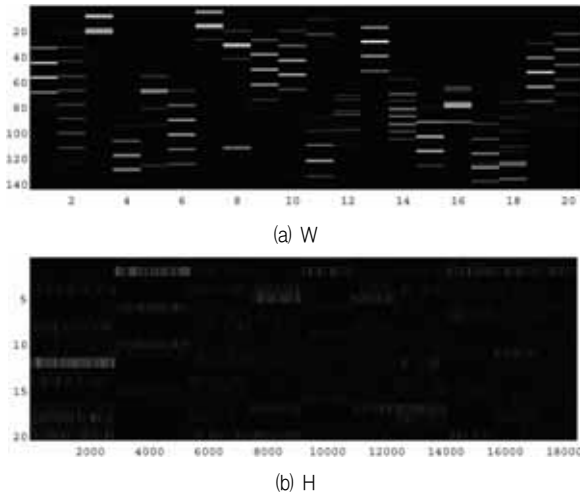
필기체 숫자 영상을 대상으로 NMF, Sparse NMF, ICA 학습을 실시해보았다. 필기체 숫자 영상은 CEDAR 데이터 베이스의 18468개이며, 각 숫자의 크기는  $12 \times 12$ 이다. 그림 1은 숫자 영상의 예들이다.

이 숫자 영상을  $144 \times 18468$  크기의 행렬  $\mathbf{V}$ 를 구성한 후, NMF와 Sparse NMF 알고리즘의 학습을 수행하였다. 이때, rank는 20을 사용하였다. 즉,  $\mathbf{W}$ 의 크기는  $144 \times 20$ 이며,  $\mathbf{H}$ 는  $20 \times 18468$ 이 된다. 그림 2는 NMF를 사용하여 얻은  $\mathbf{W}$ 와  $\mathbf{H}$ 를 보여준다.  $\mathbf{H}$ 의 모양이 10개의 클래스별로 클러스터링된 것이 보이지만 명확하게 클래스 간에 구분이 생긴

다고 보기는 어렵다.



▶▶ 그림 1. 필기체 숫자 영상



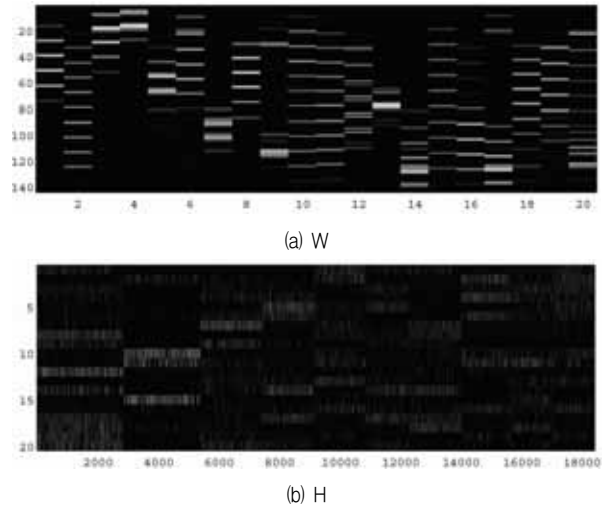
▶▶ 그림 2. NMF에 의한 한 학습에서 얻은 W와 H 행렬(rank=20)

Sparse NMF를 이용한 시뮬레이션에서는 클러스터링 효과를 얻기 위하여  $H$ 에 sparseness를 부여하기 위하여 식 (4)에서  $\eta=1$ ,  $\beta=10$  을 사용하였다. 시뮬레이션 결과 얻은  $W$ 와  $H$ 를 그림 3에 그렸다. 물론, rank=20이다. 그림 3(a)를 그림 2(a)와 비교해보면  $W$ 가 좀더 많은 영역에서 양의 값을 지니는 것을 볼 수 있다. 또한, 그림 3(b)를 그림 2(b)와 비교해보면 클러스터링 효과가 좀 더 나타났음을 알 수 있다. 그렇지만,  $H$ 가  $W$ 의 특정 basis vector만을 기반으로 클러스터링 효과가 나타난다고 명확하게 주장하기에는 부족하다.

NMF에 의해서 얻어진  $W$ 의 basis vectors는 일종의 추출된 특징으로 볼 수 있다. 이러한 특징은 ICA방법에 의해서도 얻을 수 있다.

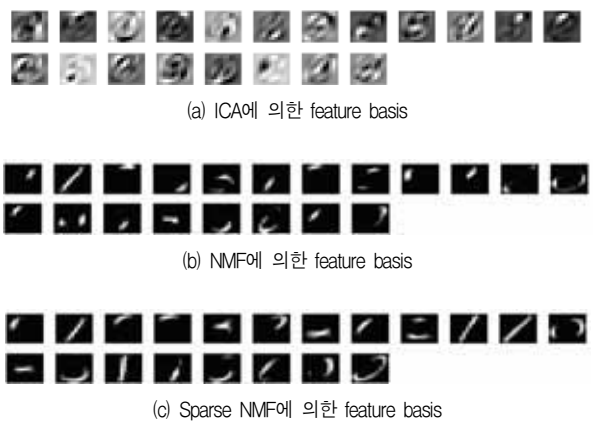
여기서는 NMF와 Sparse NMF에 의해 얻어진 feature basis들을 ICA에 의해서 얻어진 feature basis와 비교하여 보았다.

이를 위하여 동일한 필기체 숫자 데이터를 대상으로 PCA(Principal Component Analysis)를 수행하여 차원을 20으로 축소한 다음 ICA를 수행하여 feature basis를 얻었다.[4]



▶▶ 그림 3. Sparse NMF에 의한 한 학습에서 얻은 W와 H 행렬(rank=20)

그림 4는 ICA, NMF, 및 Sparse NMF에 의해서 얻어진 20개의 feature basis들을 보여준다. 그림 4(a)의 ICA에 의한 feature basis들은 global한 feature이고 그림 4(b)의 NMF에 의한 feature basis들은 아주 국소적인 local feature임을 알 수 있다. 이에 반해, Sparse NMF에 의한 결과는 그림 4(c)에서 보는 바와 같이 NMF보다는 좀더 넓은 영역에 걸쳐있는 것을 볼 수 있다.



▶▶ 그림 4. ICA, NMF, 및 Sparse NMF에 의해 얻어진 feature basis들

즉, Sparse NMF의 학습에서  $H$ 에 sparseness를 부여하는 것이  $W$ 의 sparseness를 저하시켜서 그 효과로 인하여  $W$ 의 basis들은 좀 더 넓은 영역에 걸치는 효과가 발생하였다.

#### IV. 결 론

이 논문에서는 고차원 데이터의 차원 축소관점에 feature extraction을 수행하는 NMF알고리즘과 이에 Sparseness를 부여한 Sparse NMF알고리즘의 클러스터링 효과를 필기체 숫자 영상을 대상으로 시뮬레이션하고 비교하였다. 또한, NMF와 Sparse NMF에 의해 얻어진 특징 벡터를 ICA에 의해 얻어진 것과 비교하여 보았다. 기대하였던 만큼 Sparse NMF에 의해서 클러스터링 효과가 나타나지는 않았다. 그리고, NMF는 아주 국소적인 특징을, ICA는 넓은 영역에 걸치는 특징 벡터를 추출함을 확인하였고, Sparse NMF는 NMF보다는 좀 더 넓은 영역에 걸치는 특징 벡터가 얻어지는 것을 확인하였다.

#### ■ 참고 문헌 ■

- [1] D. D. Lee and H. S. Seung, "Learning the parts of objects by non-negative matrix factorization," *Nature*, vol. 401, pp. 788-791, 1999.
- [2] H. Kim and H. Park, "Sparse non-negative matrix factorizations via alternating non-negativity-constrained least squares for microarray data analysis," *Bioinformatics*, pp. 1495-1502, 2007.
- [3] T.-W. Lee, et al., "A unifying information-theoretic framework for independent component analysis," *Computers & Mathematics with Applications*, vol. 31, no. 11, pp. 1-21, March 2000.
- [4] 오상훈, "독립성분분석에 의한 복합특징 형성," *한국콘텐츠학회 논문지*, 제3권, 제2호, pp.19-23, 2003.