

일반화된 커널 제한된 볼츠만 머신

김 동 국*

Generalized Kernel Restricted Boltzmann Machine

Dong Kook Kim*

요 약

본 논문은 커널 제한된 볼츠만 머신 (KRBM)의 일반화된 형태를 제시한다. 일반화된 KRBM은 가우시안 분포를 갖는 KRBM의 가시유닛과 은닉유닛의 불확실성을 모델링하기 위해 분산 항들을 포함한다. 일반화된 KRBM을 학습하기 위해 경사기반 contrastive divergence 알고리즘을 유도하고, 학습에 필요한 파라미터의 갱신법을 제시한다. MNIST와 STL-10 데이터를 사용한 실험에서 제안된 기법은 기존의 KRBM보다 더 향상된 복원 오차와 인식 결과를 나타낸다.

Key Words : restricted Boltzmann machine, kernel method, image classification, representation learning,

ABSTRACT

This paper presents a generalized form of kernel restricted Boltzmann machine (KRBM). The variance terms in the generalized KRBM are included to model the uncertainty of the visible and hidden units of KRBM with Gaussian distribution. The gradient-based contrastive divergence algorithm is used for the training of the generalized KRBM, and the parameter update rules are derived for the learning. Experimental results on MNIST and STL-10 dataset show that the proposed approach outperforms the conventional KRBM in terms of reconstruction error and classification accuracy.

1. 서 론

제한된 볼츠만머신(restricted Boltzmann machine, RBM)^[1-3]은 다양한 데이터로부터 검출 또는 인식을 위해 유용한 표현들을 자동적으로 추출하기 위한 비지도 표현학습(representation learning)^[4]에 널리 쓰이는 딥러닝 분야의 한 기법이다. RBM은 이미지, 음성 그리고 텍스트 등에서 좋은 특징들을 뽑아 인식성능 향상을 목적으로 주로 사용되며, 음성인식, 물체인식, 자연언어처리 등에서 널리 사용되고 있다. 또한 여러 개의 RBM을 계층적으로 쌓아서 깊은 계층을 갖는 초기 신경망을 구성하는데 이용된다^[5].

RBM은 입력 데이터에 대한 확률분포를 학습할 수 있는 에너지 기반의 단층의 생성모델 (generative model) 신경망이다^[1,2]. RBM은 일반적인 볼츠만 머신의 제한된 형태로 두 개의 유닛 그룹으로 구성된 이진 그래프(bipartite graph) 형태를 갖고 있다. 두 그룹의 유닛은 가시유닛과 은닉유닛으로 두 그룹사이의 대칭 연결을 갖고 있지만, 같은 그룹 내의 유닛사이에는 상호 연결이 없는 제한된 형태의 모델이다. 이러한 구조에 기반하여 은닉유닛들을 사용하여 가시유닛들의 고차원의 상관관계를 모델링할 수 있다. RBM은 가시유닛과 은닉유닛의 형태에 따라 다양한 구조가 제안되었다^[1]. 가장 전형적인 RBM의 구조는 입력 데이터

* 이 연구는 한국연구재단 논문연구과제(NRF-2016R1D1A1B03932813) 지원으로 수행되었습니다.

• First Author, Corresponding Author: Chonnam National University, School of Electronic and Computer Engineering, dkim@jnu.ac.kr, 정교수, 정회원

논문번호 : 202002-028-A-RE, Received February 12, 2020; Revised February 25, 2020; Accepted February 25, 2020

의 형태가 이진 값을 표현하는 가시유닛과, 이진 값을 갖는 은닉유닛으로 구성된다. 가시유닛과 은닉유닛이 모두 이진 값을 갖는 경우 Bernoulli-Bernoulli RBM (BBRBM)이라 부른다^{1,2)}. BBRBM은 이진 이미지 또는 이진 데이터를 주로 모델링하는데 사용된다. 한편 가시유닛은 실수 값을 표현하고 은닉유닛이 이진 값을 갖는 경우에는 이를 Gaussian-Bernoulli RBM(GBRBM)이라 부른다^{1,3)}. GBRBM은 주로 일반적인 입력 데이터가 음성 또는 컬러 영상들을 모델링하는데 사용된다. 입력 데이터는 평균과 분산을 파라미터로 갖는 가우시안 형태의 가시유닛으로 표현된다. 또한 은닉유닛이 특별히 ReLU(rectified linear unit) 함수를 사용하는 Noisy-ReLU RBM(NReLU RBM)⁶⁾이 제안되었는데, 은닉유닛의 샘플링을 위해 근사적인 함수가 사용되었다. 최근에 기계학습에서 널리 사용되는 커널(kernel) 개념⁷⁾과 RBM이 결합된 새로운 커널 RBM(KRBM)이 제안되었다⁸⁾. KRBM은 입력 데이터를 비선형 함수를 이용하여 고차원의 특징공간으로 암시적으로 사상하고, 그 공간에서 가우시안 분포를 갖는 가시유닛과 은닉유닛으로 구성된 구조를 갖고 있다. KRBM에서는 입력 데이터가 사상될 때 ReLU 커널함수를 사용함으로 NReLU RBM과 같이 ReLU 은닉유닛을 갖는 구조이다. KRBM은 다양한 영상인식 실험에서 기존의 RBM보다 더 나은 성능을 나타내고 있다⁸⁾.

본 논문에서는 기존의 KRBM의 일반화 형태를 제시한다. 기존 KRBM을 정의하는 에너지 함수는 가시유닛과 은닉유닛의 분산을 고려하는 항이 없어 두 유닛들을 모델링하는데 제한된다는 단점을 갖고 있다. 이러한 단점을 극복하기 위해 일반화된 KRBM은 가우시안 분포를 갖는 KRBM의 가시유닛과 은닉유닛의 불확실성을 모델링하기 위해 각각의 분산 항들을 포함한 에너지 함수를 제안한다. 그리고 일반화된 KRBM을 학습하기 위해 경사기반 CD(contrastive divergence) 학습 알고리즘을 유도하고, 학습에 필요한 일반화된 KRBM의 파라미터 갱신법을 제시한다. MNIST⁹⁾와 STL-10¹⁰⁾ 데이터 베이스를 사용하여 제안된 기법이 기존의 KRBM보다 더 향상된 복원 오차(reconstruction error)와 인식결과를 나타낸다.

본 논문의 본론 II장에서는 기존의 KRBM을 간단히 소개하고, 일반화된 KRBM 구조와 학습 알고리즘을 제시한다. III장에서는 실험과 결과에 대해 나타내고, IV에서는 결론을 맺는다.

II. 본 론

2.1 KRBM

이 단원에서 기존의 KRBM⁸⁾의 구조와 학습 방법에 대해 간단히 살펴본다. 먼저 $\mathbf{v} = (v_1, \dots, v_n)^T$ 을 원래의 n 차원 데이터 공간에서 정의된 가시변수 벡터라 하자. 여기서 T 는 전치(transpose)를 나타낸다. 이때 KRBM은 비선형함수 $\phi : R^n \rightarrow R^f$ 을 사용해 n 차원의 입력공간에서 f 차원의 특징벡터 공간으로 사상한다. $\phi(\mathbf{v})$ 는 비선형적으로 사상되는 f 차원의 특징벡터 공간에서 가시변수이다. 비록 $\phi(\mathbf{v})$ 의 형태는 분명하게 규정되지 않지만 커널트릭(kernel trick)¹⁷⁾에 근거하여 특징벡터 공간에서 두 벡터의 내적을 커널함수(kernel function) $k(\cdot, \cdot)$ 을 사용해 $k(\mathbf{v}_i, \mathbf{v}_j) = \phi(\mathbf{v}_i)^T \phi(\mathbf{v}_j)$ 로 구할 수 있다. KRBM 구조는 $\phi(\mathbf{v})$ 에 의한 가시유닛 층과 R^m 공간에서 정의된 실수 값을 갖는 m 차원 벡터 $\mathbf{h} = (h_1, \dots, h_m)^T$ 에 의한 은닉유닛 층으로 구성된다. 두 층사이의 연결 가중치를 위해 m 개의 n 차원의 가중치 벡터 $(\mathbf{w}_1, \dots, \mathbf{w}_m)$ 들을 정의하고, 이를 비선형 함수에 의해 사상된 가중치 행렬, $\Phi(W) = (\phi(\mathbf{w}_1), \dots, \phi(\mathbf{w}_m))$ 을 구성한다. 그러면 $\phi(\mathbf{w}_j)$ 는 j 번째 은닉유닛 h_j 와 특징공간에서 가시벡터 $\phi(\mathbf{v})$ 을 연결하는 가중치 벡터가 된다. 이때 KRBM은 다음과 같은 에너지 함수를 갖는다⁸⁾.

$$\begin{aligned} E(\phi(\mathbf{v}), \mathbf{h}) &= \phi(\mathbf{v})^T \phi(\mathbf{v}) - 2\mathbf{h}^T \Phi(W)^T \phi(\mathbf{v}) + \mathbf{h}^T \mathbf{h} \\ &= k(\mathbf{v}, \mathbf{v}) - 2 \sum_{j=1}^m h_j k(\mathbf{w}_j, \mathbf{v}) + \sum_{j=1}^m h_j^2 \end{aligned} \quad (1)$$

여기서 커널트릭에 의해 $k(\mathbf{v}, \mathbf{v}) = \phi(\mathbf{v})^T \phi(\mathbf{v})$ 이고, $k(\mathbf{w}_j, \mathbf{v}) = \phi(\mathbf{w}_j)^T \phi(\mathbf{v})$ 이다. 이에 근거하여 KRBM의 $\phi(\mathbf{v})$ 와 \mathbf{h} 에 대한 결합(joint) 확률분포는

$$p(\phi(\mathbf{v}), \mathbf{h}) = e^{-\frac{1}{2}E(\phi(\mathbf{v}), \mathbf{h})} / Z \text{이다. 여기서 } Z \text{는}$$

$$\text{파티션 함수, } Z = \int \int e^{-\frac{1}{2}E(\phi(\mathbf{v}), \mathbf{h})} d\mathbf{h} d\phi(\mathbf{v}) \text{ 이}$$

다. 위의 에너지 함수 하에서 각각 $\phi(\mathbf{v})$ 와 \mathbf{h} 에 대한 조건부 확률분포는 다음과 같이 다변수 가우시안(multivariate Gaussian) 형태를 갖는다⁸⁾.

$$p(\mathbf{h}|\phi(\mathbf{v})) = N(\mathbf{h}|\Phi(W)^T\phi(\mathbf{v}), I_m) \quad (2)$$

$$p(\phi(\mathbf{v})|\mathbf{h}) = N(\phi(\mathbf{v})|\Phi(W)\mathbf{h}, I_f) \quad (3)$$

여기서 I_m 와 I_f 은 각각 m 차원과 f 차원에서 단위행렬을 나타내고, $N(\mathbf{x}|\mu, \Sigma)$ 는 평균 μ , 공분산 Σ 을 갖는 확률변수 \mathbf{x} 에 대한 가우시안 분포를 나타낸다. 여기서 각각의 조건부 확률의 공분산이 단위행렬을 갖고 있기 때문에 $\phi(\mathbf{v})$ 와 \mathbf{h} 에 대해 조건부 확률을 모델링하는데 제한된 형태를 갖는 단점을 갖고 있다.

KRBM을 학습하기 위한 방법은 로그 유사도(log-likelihood) 함수를 최대화하는 경사 상승법(gradient ascent) 기반 CD 알고리즘을 이용한다. 이는 유사도를 근사화하여 내부적으로 깃스샘플링 과정을 수행한다. KRBM에서는 커널함수로 ReLU 함수를 사용한다. n 차원의 두 벡터 \mathbf{w} 와 \mathbf{v} 에 대해 ReLU 커널함수의 정의는 다음과 같다^{6,8)}.

$$k(\mathbf{w}, \mathbf{v}) = \max(\mathbf{w}^T\mathbf{v} + b, 0) \quad (4)$$

여기서 \mathbf{w} 는 가중치(weight), b 는 바이어스(bias) 성분을 나타낸다. 이러한 커널함수에 대해 1단계(single-step) CD (CD-1) 알고리즘²⁾에 의해 가중치와 bias 파라미터에 대한 유사도 함수의 경사치(gradient)에 기반한 학습 알고리즘이 제시되었다⁸⁾.

2.2 일반화된 KRBM

이 단원에서는 KRBM의 단점을 극복하기 위한 일반화된 KRBM을 제시한다. KRBM의 에너지 함수에 근거하여 구한 $\phi(\mathbf{v})$ 와 \mathbf{h} 에 대한 조건부 확률분포, 식 (2)와 (3)의 공분산 항들이 단위행렬을 갖기 때문에 가시유닛과 은닉유닛의 불확실성을 정확히 모델링하는데 한계가 존재한다. 이러한 문제점을 개선하기 위해 일반화된 KRBM에서는 다음과 같은 에너지 함수를 제안한다.

$$\begin{aligned} E(\phi(\mathbf{v}), \mathbf{h}) &= \frac{1}{\sigma^2} \phi(\mathbf{v})^T \phi(\mathbf{v}) \\ &\quad - \frac{2\mathbf{h}^T}{\sigma^2} \Phi(W)^T \phi(\mathbf{v}) + \mathbf{h}^T \Lambda^{-1} \mathbf{h} \quad (5) \\ &= \frac{1}{\sigma^2} k(\mathbf{v}, \mathbf{v}) - \frac{2}{\sigma^2} \sum_{j=1}^m h_j k(\mathbf{w}_j, \mathbf{v}) + \sum_{j=1}^m \frac{h_j^2}{\lambda_j} \end{aligned}$$

여기서 σ^2 은 $\phi(\mathbf{v})$ 에 대해, 그리고 Λ 은 \mathbf{h} 에 대한 공분산 행렬을 모델링하기 위해 추가되었다. 여기서 공분산 행렬, $\Lambda = \text{diag}(\lambda_1, \dots, \lambda_m)$ 은 대각(diagonal) 행렬이라 가정한다. 만약 $\sigma^2 = 1$ 이고, $\Lambda = I_m$ 이면 위 식 (5)의 에너지 함수는 KRBM이 되기 때문에 KRBM은 일반화된 KRBM의 특수한 경우에 해당함을 알 수 있다. 이러한 일반화된 KRBM의 에너지 함수 하에서 $\phi(\mathbf{v})$ 와 \mathbf{h} 에 대한 조건부 확률분포 식은 각각 다음과 같이 주어진다.

$$p(\mathbf{h}|\phi(\mathbf{v})) = N(\mathbf{h}|\frac{\Lambda}{\sigma^2}\Phi(W)^T\phi(\mathbf{v}), \Lambda) \quad (6)$$

$$p(\phi(\mathbf{v})|\mathbf{h}) = N(\phi(\mathbf{v})|\Phi(W)\mathbf{h}, \sigma^2 I_f) \quad (7)$$

식 (2)와 (3)과 비교하면, 일반화된 KRBM의 조건부 확률분포의 공분산 항들이 각각 Λ 와 $\sigma^2 I_f$ 에 의해 주어지고, \mathbf{h} 에 대한 평균도 공분산 항에 의해 변화됨을 알 수 있다. 따라서 샘플링 과정 중에 데이터를 공분산 항에 의해 좀 더 정확하게 샘플링할 수 있다.

일반화된 KRBM의 학습은 KRBM과 같이 ReLU 커널 함수를 사용한 CD-1 학습 알고리즘을 사용한다. 즉 로그 유사도 함수를 최대화하는 경사 상승법을 이용한다. 고차원의 특징공간에서 $\phi(\mathbf{v})$ 에 대한 로그 유사도 함수는 다음과 같다^{2,8)}.

$$\begin{aligned} l &= \ln p(\phi(\mathbf{v})|\theta) = \ln \frac{1}{Z} \int e^{-\frac{1}{2}E(\phi(\mathbf{v}), \mathbf{h})} d\mathbf{h} \\ &= \ln \int e^{-\frac{1}{2}E(\phi(\mathbf{v}), \mathbf{h})} d\mathbf{h} \quad (8) \\ &\quad - \ln \int \int e^{-\frac{1}{2}E(\phi(\mathbf{v}), \mathbf{h})} d\mathbf{h} d\phi(\mathbf{v}) \end{aligned}$$

여기서 θ 는 일반화된 KRBM의 파라미터이며, $\theta = (\mathbf{w}_{ij}, b_j, \lambda_j, \sigma^2)_{i=1\dots n, j=1\dots m}$ 로 주어진다. CD-1 알고리즘을 사용한 파라미터에 대한 로그 유사도 경사값은 다음과 같이 근사화 된다^{2,8)}.

$$\begin{aligned} \frac{\partial l}{\partial \theta} &\approx - \int p(\mathbf{h}|\phi(\mathbf{v}^{(0)})) \frac{\partial E(\phi(\mathbf{v}^{(0)}), \mathbf{h})}{\partial \theta} d\mathbf{h} \\ &\quad + \int p(\mathbf{h}|\phi(\mathbf{v}^{(1)})) \frac{\partial E(\phi(\mathbf{v}^{(1)}), \mathbf{h})}{\partial \theta} d\mathbf{h} \quad (9) \end{aligned}$$

여기서 두 번째 항의 기댓값은 모델 분포로부터 깃스 샘플링을 통해 얻어진다. 깃스 샘플링은 입력 데이터 샘플을 $\mathbf{v}^{(0)}$ 으로 초기화하고 $\phi(\mathbf{v}^{(0)})$ 으로 사상한다. 그리고 식 (6)의 $p(\mathbf{h}|\phi(\mathbf{v}^{(0)}))$ 로부터 $\mathbf{h}^{(0)}$ 을 샘플링 한다. 순차적으로 식 (7)의 $p(\phi(\mathbf{v})|\mathbf{h}^{(0)})$ 으로부터 $\mathbf{v}^{(1)}$ 의 샘플을 얻는다⁸⁾. 이러한 깃스 샘플링을 통해 일반화된 KRBM의 파라미터에 대한 경사값을 식 (9)을 통해 얻을 수 있다. 다만 Λ 와 σ^2 은 각각 공분산 항을 나타내기 때문에 항상 양수 값을 갖는 제약 조건이 존재한다. 이를 위해 두 파라미터에 대해 로그 값을 취한 변수 치환을 $\alpha^2 = \ln \sigma^{-2}$, $\sigma^2 = e^{-\alpha^2}$ 와 $\beta_j = \ln \lambda_j^{-1}$, $\lambda_j = e^{-\beta_j}$ 로 수행하고, α^2 와 β_j 에 대한 경사값을 구하게 되면 자연스럽게 두 파라미터가 양의 값을 갖게 된다. ReLU 커널함수를 사용하는 경우, 일반화된 KRBM의 파라미터인 가중치 w_{ij} , 바이어스 b_j 그리고 α^2 와 β_j 에 대한 경사값은 식 (9)로부터 다음과 같이 쉽게 유도할 수 있다.

$$\nabla w_{ij} = v_i^{(0)} \hat{h}_j^{(0)} - v_i^{(1)} \hat{h}_j^{(1)} \quad (10)$$

$$\nabla b_j = \hat{h}_j^{(0)} - \hat{h}_j^{(1)} \quad (11)$$

$$\begin{aligned} \nabla \alpha^2 &= e^{-\alpha^2} (k(\mathbf{v}^{(1)}, \mathbf{v}^{(1)}) - k(\mathbf{v}^{(0)}, \mathbf{v}^{(0)})) \\ &+ 2 \sum_{j=1}^m \hat{h}_j^{(0)} k(\mathbf{w}_j, \mathbf{v}^{(0)}) - 2 \sum_{j=1}^m \hat{h}_j^{(1)} k(\mathbf{w}_j, \mathbf{v}^{(1)}) \end{aligned} \quad (12)$$

$$\nabla \beta_j = e^{-\beta_j} (\hat{h}_j^{2,(1)} - \hat{h}_j^{2,(0)}) \quad (13)$$

여기서 $\hat{h}_j^{(k)}$ 와 $\hat{h}_j^{2,(k)}$, $k=0,1$ 는 식 (5)와 식 (6)의 조건부 확률분포, $p(\mathbf{h}|\phi(\mathbf{v}^{(k)}))$ 로부터 구한 h_j 의 조건부 1차 및 2차 모멘트(moment)을 나타내며, 각각은 다음과 같이 주어진다.

$$\begin{aligned} \hat{h}_j^{(k)} &= E[h_j|\phi(\mathbf{v}^{(k)})] \\ &= \frac{\lambda_j}{\sigma^2} \max(\mathbf{w}_j^T \mathbf{v}^{(k)} + b_j, 0) \end{aligned} \quad (14)$$

$$\hat{h}_j^{2,(k)} = E[h_j^2|\phi(\mathbf{v}^{(k)})] = \lambda_j + (\hat{h}_j^{(k)})^2 \quad (15)$$

여기서 $E[|]$ 는 조건부 평균을 나타낸다. 최종적으로 일반화된 KRBM의 학습 알고리즘을 요약하면 다음과 같다.

일반화된 KRBM 학습 알고리즘(CD-1 기반)
 입력 : 학습 batch S , 학습률 η
 출력 : 학습된 파라미터

$$\theta = (w_{ij}, b_j, \lambda_j, \sigma^2)_{i=1\dots n, j=1\dots m}$$

1. θ 값 초기화
2. for all the $\mathbf{v} \in S$ do
3. $\mathbf{v}^{(0)} \leftarrow \mathbf{v}$
4. 식 (6), $p(\mathbf{h}|\phi(\mathbf{v}^{(0)}))$ 을 사용해 $\mathbf{h}^{(0)}$ 샘플링
5. 식 (7), $p(\phi(\mathbf{v})|\mathbf{h}^{(0)})$ 을 사용해 $\mathbf{v}^{(1)}$ 샘플링
6. 식 (14), (15)을 사용해 $(\hat{h}_j^{(k)}, \hat{h}_j^{2,(k)})_{k=0,1}$ 계산
7. 식 (10) - (13)을 사용해 파라미터 갱신
 $\theta \leftarrow \theta + \eta \nabla \theta$

III. 실험 및 결과

제안된 일반화된 KRBM(GKRBM)의 성능을 평가하기 위해 MNIST⁹⁾와 ST-10¹⁰⁾ 데이터에 대해 비지도 특징학습 및 인식 실험을 수행하였다. 학습을 위해 CD-1 알고리즘을 사용하였고, 경사 승강법의 학습률은 MNIST는 0.001, STL-10는 0.0001을 사용하고, 0.9의 모멘텀을 사용하였다. 가중치의 초기값은 영평균과 0.01의 표준편차를 갖는 가우시안으로부터 발생된 랜덤값을 사용하였다. GKRBM에서 $\sigma^2 = 1$, $\Lambda = I_m$ 으로 초기값을 설정하였다. 학습을 위한 배치 크기는 100이며, 200 epoch 만큼 학습하였다. 은닉유닛의 수는 가변적으로 사용하여 성능을 비교하였다.

3.1 MNIST

MNIST는 손으로 쓴 숫자 0-9 사이의 28×28 크기의 grayscale 형식의 영상이다⁹⁾. 60,000개의 학습 샘플과 10,000개의 테스트 샘플로 구성된다. 입력은 $28 \times 28 = 784$ 크기의 벡터이다.

첫 번째 실험에서 1024개의 은닉유닛을 갖는 KRBM과 GKRBM을 학습하였다. 그림 1은 epoch에 따른 평균 복원 오차를 나타낸다. 그림에서 보듯이

1) 평균 복원 오차는 CD-1알고리즘에서 데이터 샘플 N개에 대

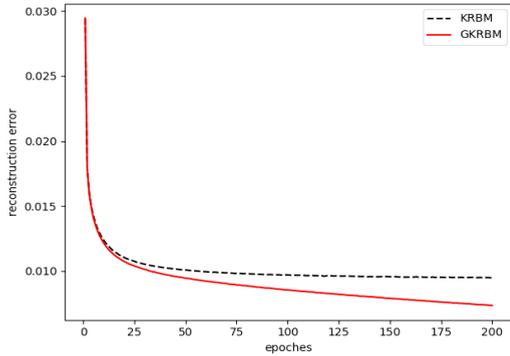


그림 1. MNIST에서 1024개 은닉유닛을 갖는 KRBM과 일반화된 KRBM 복원 오차
Fig. 1. Reconstruction errors of KRBM and GKRBM with 1024 hidden units on MNIST

제한된 GKRBM은 기존 KRBM에 비해 더 낮은 복원 오차를 갖고 수렴하는 것을 알 수 있다. 학습 데이터에 대해 평균적으로 KRBM은 0.0095을, GKRBM은 더 낮은 0.0074의 평균 복원 오차 값을 나타냈다. 그림 2은 원 영상, KRBM과 GKRBM을 사용해 복원된 영상을 나타낸다. 영상 위의 숫자는 각 영상에 대한 복원 오차 값을 나타낸다. 그림 2에서 보듯이 복원된 영상들은 원래의 영상의 흐릿해진(blurred) 영상을 띄고 있으며 두 기법 모두 전체적인 모양을 잘 복원하고 있다. 모든 숫자 영상에 대해 GKRBM이 더 낮은 복원 오차를 갖기 때문에 KRBM보다 더 잘 복원함을 알 수 있다. 그리고 각각 알고리즘의 특징 표현에 대한 시각화를 살펴보기 위해 학습된 필터(가중치)에 대한 값을 그림으로 표현하였다. 그림 3은 KRBM와 GKRBM은 1024개의 필터 중에 분산이 가장 큰 100개에 대한 그림이다. 일반적으로 RBM은 Gabor 필터와 닮은 국부적인 형태의 필터를 나타낸 것으로 잘 알려졌다^{2,31}. 그림에서 보듯이 두 필터 모두 비슷한 모양을 갖고 있지만, GKRBM의 필터가 KRBM보다 조금 더 sparse한 형태를 갖는 것을 알 수 있다.

두 번째 실험에서는 표현학습을 통해 MNIST 인식 실험을 수행하였다. 이를 위해 입력이 주어진 경우, 학습된 RBM을 사용하여 은닉유닛의 수만큼 새로운 특징벡터로 추출하고, softmax 인식을 사용해서 인식 실험을 수행하였다. 그림 4는 가변적인 은닉유닛의 수에 따른 테스트 데이터에 대한 인식 정확도를 나타낸다. 비교를 위해 NReLU RBM의 성능도 표시하였다.

해 원영상 $v^{(0)}$ 와 샘플링된 영상 $v^{(1)}$ 사이의 거리의 평균, $\frac{1}{N} \sum_{k=1}^N \|v_k^{(0)} - v_k^{(1)}\|^2$ 으로 정의된다.

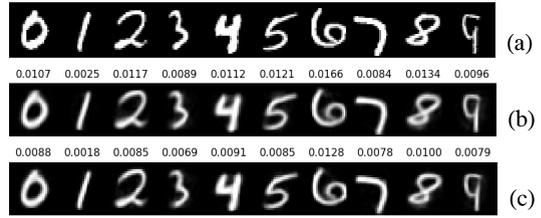


그림 2. MNIST에서 원 영상과 KRBM과 GKRBM을 사용한 복원된 영상. (a) 원 영상 (b) KRBM, (c) GKRBM
Fig. 2. Original image and reconstruction images using KRBM and GKRBM on MNIST.

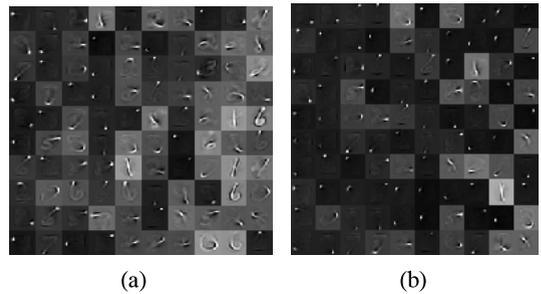


그림 3. MNIST에서 학습된 1024개 은닉유닛을 가지는 RBM에서의 가중치의 시각화. (a) KRBM, (b) GKRBM
Fig. 3. Visualization of the weights of RBMs with 1024 hidden units trained on MNIST.

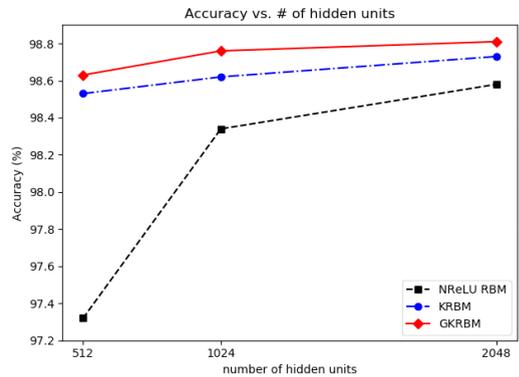


그림 4. MNIST에서 가변적인 은닉유닛 수에 따른 테스트 인식 정확도
Fig. 4. Test classification accuracy with various number of hidden units on MNIST

전체적으로 은닉유닛 수가 증가할수록 인식성능이 향상됨을 알 수 있다. 그리고 GKRBM이 KRBM보다 서로 다른 은닉유닛 수에서 약간 더 향상된 성능을 나타내었다. 따라서 제안된 GKRBM 기법에서 가시유닛과 은닉유닛의 공분산을 고려해서 모델링하는 것이 기존 기법에 비해 비지도 특징학습에 효과적임을 알 수 있다.

3.2 STL-10

STL-10^[10]은 특별히 비교사 특징학습을 위해 10개의 클래스로 구성되며, 96×96 픽셀 크기의 컬러 영상이다. 비교사 학습을 위한 데이터 수는 100,000개이며, 교사학습을 위해 클래스당 500개의 학습 영상과 800개 테스트 영상으로 구성된다. STL-10 데이터를 비지도 학습과 인식에 사용하기 위해 MNIST와 다르게 전처리 과정을 수행한다. 본 논문에서는 비지도 특징학습에 널리 쓰이는 방법을 따른다^[10]. 먼저 계산량과 메모리 사용을 줄이기 위해 96×96 크기의 영상을 32×32 크기로 줄인다. 그리고 학습을 위해 6×6 크기의 컬러 패치(patch)를 클래스가 없는 100,000개 영상의 임의의 위치에서 총 500,000개를 추출하여 사용한다. 따라서 입력의 개수는 $6 \times 6 \times 3 = 108$ 이다. 모든 데이터는 whitening을 과정을 수행하고, KRBM과 GKRBM에 대해 영평균을 갖도록 정규화 한다.

표 1은 1024개 은닉유닛을 갖는 경우의 KRBM과 GKRBM의 평균 복원 오차를 나타낸다. MNIST와 같이 GKRBM이 더 낮은 복원 오차를 갖는 것을 알 수 있다. 그림 5는 6×6 크기의 가중치 중에서 분산이 가장 큰 100개에 대한 시각화를 나타낸다. KRBM과 GKRBM 모두 컬러 영상의 edge 등의 특징을 추출할 수 있는 형태를 가지고 있다. MNIST와 마찬가지로 GKRBM이 KRBM보다 더 sparse한 형태를 갖고 특징들을 추출함을 알 수 있다.

STL-10의 경우에 인식 실험을 수행하기 위해 입력 영상에 대해 같은 간격으로 패치들을 생성하고 RBM을 통해 특징들을 추출한다^[10]. 최종적으로 한 입력 영상당 가변적인 은닉유닛의 수에 따라 특징 벡터를 발생하였다. 이를 MNIST와 마찬가지로 softmax 인식기를 사용해 인식실험을 수행하였다. 그림 6은 가변적인 은닉유닛의 수에 따른 테스트 데이터에 대한 인식 정확도를 나타낸다. 특징학습이 없는 경우 31.8%의 매우 낮은 인식률을 나타내었다. 여기에서도 GKRBM이 KRBM에 약간 더 높은 인식률을 나타내었다. 따라서 STL-10 데이터에서도 제안된 GKRBM이 효과적임을 알 수 있다.

표 1. STL-10에서 1024개 은닉유닛을 갖는 KRBM과 GKRBM의 복원 오차
Table 1. Reconstruction errors of KRBM and GKRBM with 1024 hidden units trained on STL-10.

	KRBM	GKRBM
평균 reconstruction error	0.0897	0.0813

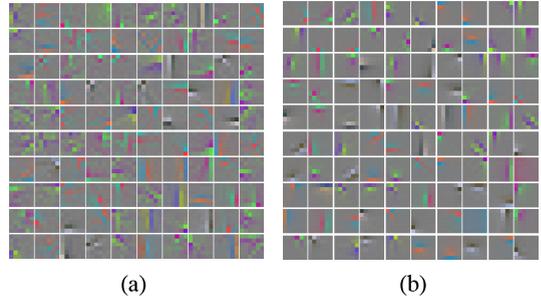


그림 5. STL-10에서 학습된 1024개 은닉유닛을 갖는 6×6 크기의 가중치(필터)의 시각화. (a) KRBM, (b) GKRBM
Fig. 5. Visualization of the weights of 6×6 size with 1024 hidden units trained on STL-10.

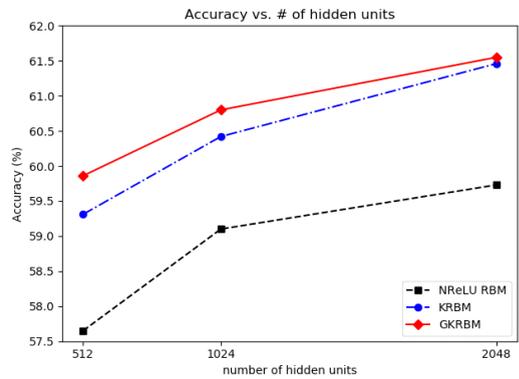


그림 6. STL-10에서 가변적인 은닉유닛 수에 따른 테스트 인식 정확도
Fig. 6. Test classification accuracy with various number of hidden units on STL-10

IV. 결론

본 논문에서는 가우시안 분포를 갖는 KRBM의 가시유닛과 은닉유닛의 불확실성을 모델링하기 위해 분산 항들을 포함하여 에너지 함수를 갖는 일반화된 KRBM을 제안하였다. 일반화된 KRBM의 파라미터 학습을 위해 CD-1에 기반한 알고리즘을 제시하였다. MNIST와 STL-10 데이터에서 학습과 인식 실험결과 제안된 KRBM은 기존의 KRBM보다 더 낮은 복원 오차와 더 높은 인식성능을 나타내었다.

References

[1] G. Hinton, "A practical guide to training restricted Boltzmann machines," *Tech. Rep. UTMIL TR 2010 - 003*, University of Toronto, 2010.
[2] A. Fischer and C. Igel, "Training restricted

Boltzmann machines: An introduction,” *Pattern Recognition*, vol. 47, pp. 25-39, 2014.

[3] K. Cho, A. Ilin, and T. Raiko, “Improved learning of Gaussian-Bernoulli restricted Boltzmann machines,” in *Proc. 21st ICANN*, pp. 10-17, 2011.

[4] Y. Bengio, A. Courville, and P. Vincent, “Representation learning: A review and new perspectives,” *IEEE Trans. PAMI*, special issue Learning Deep Architectures, vol. 35, no. 8, pp. 1798-1828. Aug. 2013.

[5] G. E. Hinton and R. R. Salakhutdinov, “Reducing the dimensionality of data with neural networks,” *Science*, vol. 313, pp. 504-507, 2006.

[6] V. Nair and G. E. Hinton, “Rectified linear units improve restricted Boltzmann machines,” *ICML*, pp. 807-814, 2010.

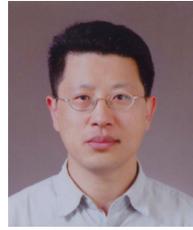
[7] B. Schölkopf and A. Smola, *Learning with Kernels*, MIT Press, 2002.

[8] D. Kim and J. W. Shin, “Kernel-based restricted Boltzmann machine for unsupervised feature learning,” *J. KICS*, vol. 44, no. 9, pp. 1633-1640, 2019.

[9] Y. LeCun, L. Bottou, Y. Bengio, and P. Haffner, “Gradient-based learning applied to document recognition,” in *Proc. IEEE*, vol. 86, no. 11, pp. 2278-2324, Nov. 1998.

[10] A. Coates, H. Lee, and A. Ng, “An analysis of single-layer networks in unsupervised feature learning,” in *Proc. Fourteenth Int. Conf. Artificial Intell. and Statistics*, PMLR 15, pp. 215-223, 2011.

김 동 국 (Dong Kook Kim)



1989년 2월 : 전남대학교 전자공학과 학사

1991년 2월 : 포항공과대학 전자전기공학과 석사

2003년 2월 : 서울대학교 전기컴퓨터공학부 박사

1991년 2월~1999년 2월 : 삼성 전자 전문연구원

2003년 4월~2004년 2월 : 한국전자통신연구원 선임연구원

2004년 2월~현재 : 전남대학교 전자컴퓨터공학부 교수 <관심분야> 딥러닝, 기계학습, 음성처리

[ORCID: 0000-0001-9316-7069]