

비지도 특징학습을 위한 새로운 에너지 기반 은닉변수 모델

곽 봉*, 김 동 국^o

A New Energy-Based Latent-Variable Model for Unsupervised Feature Learning

Guo Peng*, Dong Kook Kim^o

요 약

본 논문은 비지도 특징학습을 위한 새로운 에너지기반 은닉변수 모델(EBLVM)을 제안한다. EBLVM의 결합 확률밀도함수는 심층 신경망에 의해 변환된 연속적인 가시변수와 은닉변수에 대한 새로운 에너지 함수를 정의한다. 경사도 기반 contrastive divergence 알고리즘을 사용하여 새로운 EBLVM의 파라미터를 훈련한다. EBLVM은 심층구조를 갖고 모든 은닉층을 연합하여 학습하기 때문에 각 층에서 특징학습을 위한 효과적인 특징들을 추출할 수 있다. Fashion MNIST와 CIFAR10 데이터를 사용한 비교사 특징학습 실험에서 제안된 기법은 기존의 stacked RBM, DBN, DBM 그리고 DEM보다 더 향상된 인식성능을 나타낸다.

Key Words : unsupervised feature learning, latent variable model, energy-based model, restricted Boltzmann machine, image classification

ABSTRACT

This paper proposes a new energy-based latent-variable model (EBLVM) for unsupervised feature learning. The joint probability density function of EBLVM defines a new energy function for the continuous visible and hidden variables in which the visible variable is transformed by the deep neural network. We train the parameters of a new EBLVM using a gradient-based contrastive divergence algorithm. Since EBLVM has a deep structure and learns by combining all hidden layers, effective features for feature learning can be extracted from each layer. In comparative feature learning experiments using Fashion MNIST and CIFAR10 data, the proposed method shows better recognition performance than the existing stacked RBM, DBN, DBM, and DEM.

I. 서 론

최근 비지도 특징학습(unsupervised feature learning)은 기계학습 가운데 중요한 분야이며, 레이블(label)이 없는 많은 데이터로부터 여러 가지 유용된

특징을 자동으로 추출하기 위한 기법이다. 추출된 특징은 여러 가지 데이터의 분석 및 인식 등의 응용 분야에 적용된다. 이러한 특징학습을 위한 다양한 모델과 학습 기법들이 개발되었으며, 영상^[1], 음성^[2], 텍스트^[3] 등 분야에 널리 사용되고 있다. 비지도 특징학습

* First Author : Chonnam National University, Department of Electronic Engineering, bongkua0379@gmail.com, 학생회원
^o Corresponding Author : Chonnam National University, Department of Electronic Engineering, dkim@jnu.ac.kr, 종신회원
 논문번호 : 202303-043-A-RU, Received March 7, 2023; Revised March 31, 2023; Accepted April 4, 2023

과 관련된 많은 기법중에 에너지기반 모델 (EBM, Energy-based Model)^[4-10]에 대한 연구가 활발하게 진행되고 있다.

EBM (또는 비정규화 확률모델)은 에너지 함수를 통해 관측 데이터의 확률분포를 모델링하기 위해 사용되는 기법이다^[4]. EBM은 다른 모델에 비해 확률분포 형태에서 훨씬 덜 제약된 함수 형태를 갖고 있다. 정규화된 확률분포 대신에 에너지 함수만을 통해 데이터의 분포를 모델링하는 비정규화된 형태를 갖고 있다. 이러한 특징들 때문에 다양한 데이터를 유용하게 모델링할 수 있고, 심층 신경망과 결합된 형태로 사용되어 최근에 생성모델로서 주목받고 있다^[9,10]. 그러나 비정규화 상수로 인하여 정확한 유사도 계산과 이 모델로부터 데이터들의 정확한 생성이 어려운 단점들을 갖고 있다. 이러한 것들은 EBM의 학습을 어렵게 만드는 요인이 되고 있다.

최근에 많은 EBM의 기법들이 개발되었다. 그 중에서 가장 많이 사용되는 형태는 에너지기반 은닉변수 모델 (EBLVM, Energy-based Latent-Variable Model)^[4,5,10]이다. EBLVM는 EBM의 하나의 형태로, 가시변수와 은닉변수간의 에너지 함수를 정의하여 결합 확률분포를 모델링 하는 기법이다. 많은 EBLVM중에서 가장 대표적인 것은 RBM(Restricted Boltzmann Machine)^[11-14]이다. RBM은 입력 데이터에 대한 확률분포를 학습할 수 있는 에너지 기반의 단층의 생성모델 신경망이다^[11]. RBM은 음성, 영상, 텍스트 영역에서 비지도 특징 학습 방법으로 우수한 성과를 거두었다. RBM의 형태는 입력과 은닉변수에 따라 다양한 형태를 갖고 있다. 또한 심층구조의 EBLVM으로 여러 개의 RBM을 계층적으로 쌓아 구성한 DBN(Deep Belief Network)^[15], DBM(Deep Boltzmann Machine)^[16], DEM(Deep Energy Model)^[6] 그리고 최근의 CEBM(Conjugate Energy-Based Model)^[10] 등이 개발되어 특징학습 또는 생성모델에 이용되고 있다.

본 논문에서는 비지도 특징학습을 위한 새로운 EBLVM을 제시한다. 새로운 EBLVM은 deterministic 은닉변수로 구성된 하위층(심층신경망)과 마지막 층은 stochastic 은닉변수로 구성된다. 이러한 구조하에 제안된 모델의 핵심 아이디어는 가시벡터와 은닉변수의 결합 확률밀도함수를 정의하기 위해 하위층의 심층 신경망에 의해 변환된 가시변수와 마지막 은닉변수에 대해 새로운 에너지 함수를 모델링하는 것이다. 심층 구조가 갖고 모든 은닉층들을 연합하여 학습하기 때문에 각 층마다 특징학습을 위한 유용한 특징을 잘 추출할 수 있다. 새로운 EBLVM을

학습하기 위해 경사기반 contrastive divergence (CD)^[11-15]알고리즘을 사용하여 파라미터를 갱신한다. Fashion MNIST와 CIFAR10 데이터를 사용한 비지도 특징학습을 통한 이미지 인식 실험에서 제안된 기법은 기존의 Stacked RBM, DBN, DBM 그리고 DEM보다 더 향상된 인식성능을 나타낸다.

본 논문은 II장 본문에서는 기존의 EBM을 간단히 소개하고, 새로운 EBLVM의 구조와 학습 알고리즘을 제시한다. III장에서는 실험과 결과에 대해 나타내고, IV장에서는 결론을 맺는다.

II. 에너지기반 은닉변수 모델

2.1 EBM

이 단원에서 기존의 EBM에 대해 간단히 살펴본다^[4-6]. EBM은 생성모델 가운데 하나의 기법으로 관측 데이터에 대한 확률분포를 모델링하는데 주로 사용된다. 알려지지 않은 데이터 분포 $p_{data}(\mathbf{x})$ 로부터 추출된 i.i.d. 데이터 셋, $\{\mathbf{x}^{(i)}\}_{i=1}^N$ 가 주어진 경우, EBM은 데이터 분포 $p_{data}(\mathbf{x})$ 을 근사화하기 위해 파라미터 θ 를 갖는 모델 분포 $p_{\theta}(\mathbf{x})$ 를 사용하여 학습하는 모델이다. EBM은 Gibbs 분포를 통해 가시변수 $\mathbf{x} \in R^n$ 의 확률 밀도 함수를 $p_{\theta}(\mathbf{x}) = \frac{\exp\{-E_{\theta}(\mathbf{x})\}}{Z_{\theta}}$, $Z_{\theta} = \int \exp\{-E_{\theta}(\mathbf{x})\} d\mathbf{x}$ 으로 정의한다. 여기 E_{θ} 는 에너지 함수이고, Z_{θ} 는 정규화 상수, 또는 파티션(partition) 함수로 부른다. EBM의 목적은 학습을 통해 실제 데이터 분포 $p_{data}(\mathbf{x})$ 을 근사하는 $p_{\theta}(\mathbf{x})$ 의 파라미터 θ 를 구하는 것이다. θ 를 구하기 위해 p_{data} 와 p_{θ} 사이의 Kullback-Leibler(KL) divergence, $D_{KL}(p_{data} \parallel p_{\theta})$ 을 최소화하기 위한 경사하강법이 널리 쓰이고 있다. 이는 다음과 같이 평균 로그유사도(expected log-likelihood), $L(\theta) = -\mathbf{E}_{\mathbf{x} \sim p_{data}(\mathbf{x})}[\log p_{\theta}(\mathbf{x})]$ 을 최대화하는 것과 같다. 이때, EBM에 대한 로그 유사도에 대한 경사도는 다음과 같은 형태로 주어진다^[5].

$$\begin{aligned} \nabla_{\theta} L(\theta) &= -\mathbf{E}_{\mathbf{x} \sim p_{data}(\mathbf{x})}[\nabla_{\theta} E_{\theta}(\mathbf{x})] + \mathbf{E}_{\mathbf{x}' \sim p_{\theta}(\mathbf{x}')}[\nabla_{\theta} E_{\theta}(\mathbf{x}')] \end{aligned} \quad (1)$$

여기서 $\mathbf{E}[\cdot]$ 는 평균을 나타낸다. 식(1)의 첫 번째 항은

학습 데이터로부터 쉽게 구할 수 있지만, 두 번째 항의 경사도를 구하기 위해서는 Markov chain Monte Carlo (MCMC)와 같은 근사적인 샘플링 기법을 사용한다. EBM에서는 관측벡터에 대한 데이터를 생성하기 위해서 stochastic gradient Langevin dynamics (SGLD) 기법이 아래와 같이 가장 일반적으로 사용된다⁵⁾.

$$\mathbf{x}'_t = \mathbf{x}'_{t-1} + \frac{\alpha_t}{2} \nabla_{\mathbf{x}'} E_{\theta}(\mathbf{x}'_{t-1}) + \sqrt{\alpha_t} \epsilon_t, \epsilon_t \sim N(0, I) \quad (2)$$

여기서 $\alpha_t > 0$ 인 상수이다. 샘플이 생성되는 과정은 초기값 \mathbf{x}'_0 에서 시작하여 매 단계별로 반복적인 가산적인 가우시안 잡음을 더한 후 일련의 경사도 값의 갱신을 통해 생성하게 된다.

은닉변수(latent variable) EBM(EBLVM)은 은닉변수를 에너지 함수에 통합한 EBM의 하나의 종류로, 가시변수 $\mathbf{x} \in R^n$ 와 은닉 변수 $\mathbf{h} \in R^m$ 의 결합밀도함수를 다음과 같이 정의하는 모델이다.

$$p_{\theta}(\mathbf{x}, \mathbf{h}) = \frac{\exp\{-E_{\theta}(\mathbf{x}, \mathbf{h})\}}{Z_{\theta}} \quad (3)$$

여기서 $E_{\theta}(\mathbf{x}, \mathbf{h})$ 는 가시변수와 은닉변수에 대한 에너지 함수이다. 이 때 가시변수에 대한 marginal 확률분포 함수는 연속값을 갖는 은닉변수에 대해, $p_{\theta}(\mathbf{x}) = \frac{1}{Z_{\theta}} \int \exp\{-E_{\theta}(\mathbf{x}, \mathbf{h})\} d\mathbf{h}$ 와 같이 구할 수 있다.

EBLVM에 대해 잘 알려진 형태는 RBMs^[11-14], DBN^[15], DBM^[16] 그리고 DEM^[6] 등이 있다. RBM은 단층의 은닉층을 갖는 EBLVM으로 가시변수와 은닉변수가 갖는 값의 형태에 따라 다양하게 구분된다. 가시변수와 은닉변수 모두 이진값을 갖는 경우, 이를 BBRBM(Bernoulli-Bernoulli RBM)이라 부르며, 가시변수가 연속적인 실수 값을 갖는 경우, GBRBM(Gaussian-Bernoulli RBM)이라 부른다. 또한 비선형 함수를 통해 가시적 데이터와 가중치 벡터를 비선형 특징 공간에 암시적으로 투영한 후 커널 트릭(kernel trick)을 사용하여 커널 특징 공간에서 에너지 함수를 정의한 KRBM(Kernel RBM)^[17,18] 등이 있다.

DBN^[15]은 RBM로 구성된 심층 EBLVM이다.

DBN은 그림1 (a)와 같이 최상위 계층이 무방향(undirected) 연결이고 나머지 계층이 유향 연결(directed)인 그래프 확률 모델이다. DBN의 학습은 RBM을 계층적으로 쌓아서 layer-wise 방식으로 학습하고, 각 RBM의 입력은 이전 계층의 사후 분포(posterior distribution)로 모델링한다. 즉 현재 은닉층의 훈련이 끝나면 이 층의 출력을 다음 층의 입력으로 삼아 훈련을 계속한다.

DBM^[16]은 DBN과 유사한 심층 구조의 EBLVM의 대표적인 모델이다. DBM은 그림1 (b)와 같이 네트워크의 모든 계층이 무방향으로 연결되어 있고 학습은 layer-wise 방식으로 학습한 후 모든 층을 연합해서 훈련한다. DBN과 달리 DBM의 추리 과정은 아래로부터 위로 전달되는 것 외에 위에서 아래로 향하는 피드백도 포함하는데 이는 심층 신경망과 유사하다.

DEM^[6]은 그림 1 (c)와 같은 심층 구조의 EBLVM의 한 종류로 여러 개의 deterministic 은닉층과 하나의 stochastic 은닉층으로 구성된 형태이다. 먼저 입력을 feedforward 신경망을 통해 결정적으로 변환하고, 이 출력을 하나의 은닉층을 갖는 RBM과 같은 형태로 모델링한다. 연속적인 입력벡터 \mathbf{x} 와 마지막 은닉층이 이진값을 갖는 stochastic 은닉변수 \mathbf{h} 인 경우, EBM의 에너지 함수는 다음과 같이 정의된다⁶⁾.

$$E_{\theta}(\mathbf{x}, \mathbf{h}) = \frac{1}{\sigma^2} \mathbf{x}^T \mathbf{x} - \mathbf{h}^T W \mathbf{g}_{\lambda}(\mathbf{x}) - \mathbf{a}^T \mathbf{x} - \mathbf{b}^T \mathbf{h} \quad (4)$$

여기서 $\theta = \{\sigma, W, \mathbf{a}, \mathbf{b}\}$ 는 DEM의 파라미터이며, $\mathbf{g}_{\lambda}(\mathbf{x})$ 는 파라미터 λ 을 갖는 feedforward 신경망을 나타낸다. 만약 $\mathbf{g}_{\lambda}(\mathbf{x}) = \mathbf{x}$ 이면, DEM은 GBRBM이 되기 때문에 DEM은 심층 구조를 갖는 GBRBM의 일반적인 형태라 할 수 있다. 위의 다양한 EBLVM을 학습하기 위해 경사기반 contrastive divergence(CD) 알고리즘을 사용하여 파라미터를 갱신한다⁵⁾.

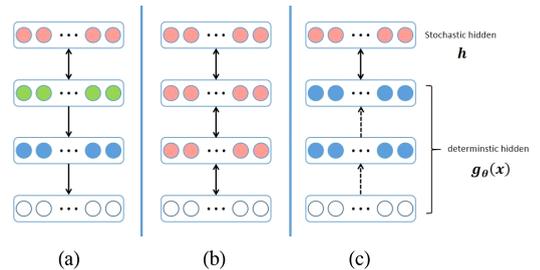


그림 1. 다양한 EBLVM의 구조⁶⁾. (a) DBN (b) DBM (c) DEM

Fig. 1. Structure of various EBLVMs^[6].

2.2 제안된 EBLVM

이 단원에서는 새로운 EBLVM을 위한 에너지 함수를 제시한다. 여기서 사용되는 EBLVM의 구조는 그림1 (c)와 같은 DEM의 형태를 갖는다. 기존의 DEM과 다르게 가시변수 \mathbf{x} 와 stochastic 은닉변수 \mathbf{h} 모두 연속적인 값을 갖는다고 가정한다. 이때 식(3)에 기반하여 본 논문에서 제안된 에너지 함수는 다음과 같다.

$$E_{\theta}(\mathbf{x}, \mathbf{h}) = \frac{1}{2}(\mathbf{x}^T \mathbf{x} - 2\mathbf{h}^T \mathbf{f}_{\theta}(\mathbf{x}) + \mathbf{h}^T \mathbf{h}) \quad (5)$$

여기서 $\mathbf{f}_{\theta}(\mathbf{x})$ 는 파라미터 θ 을 갖는 입력층에서 마지막 은닉층까지 feedforward 신경망을 나타낸다. 위 에너지 함수의 특징은 가시변수와 은닉변수 모두 각각의 2차 항들을 갖고 있다는 것이다. 이러한 특징 때문에 \mathbf{x} 와 \mathbf{h} 에 대한 각각의 조건부 확률분포는 다음과 같이 간단한 형태로 나타낸다.

$$p_{\theta}(\mathbf{h} | \mathbf{x}) = N(\mathbf{h}; \mathbf{f}_{\theta}(\mathbf{x}), \mathbf{I}) \quad (6)$$

$$p(\mathbf{x} | \mathbf{h}) = \frac{1}{Z_{\theta}(\mathbf{h})} \exp\left\{-\frac{1}{2}(\mathbf{x}^T \mathbf{x} - 2\mathbf{h}^T \mathbf{f}_{\theta}(\mathbf{x}))\right\} \quad (7)$$

여기서 $N()$ 는 Gaussian 확률분포를 나타내며, \mathbf{I} 는 단위행렬이다. 여기서 주목할 것은 $\mathbf{f}_{\theta}(\mathbf{x})$ 가 가시변수 \mathbf{x} 가 주어졌을 때 은닉변수 \mathbf{h} 의 평균에 해당한다는 것이다. 따라서 입력벡터 \mathbf{x} 가 주어진 경우, 새로운 특징 벡터는 심층 신경망 은닉층의 출력을 통해 간단히 구할 수 있다. 그리고 가시변수에 대한 marginal 확률분포도 아래와 같은 간단한 형태를 가지게 된다.

$$p_{\theta}(\mathbf{x}) = \frac{\exp\{-E_{\theta}(\mathbf{x})\}}{Z_{\theta}}, \quad E_{\theta}(\mathbf{x}) = \frac{1}{2}(\mathbf{x}^T \mathbf{x} - \mathbf{f}_{\theta}^T(\mathbf{x}) \mathbf{f}_{\theta}(\mathbf{x})) \quad (8)$$

여기서 $E_{\theta}(\mathbf{x})$ 는 marginal 확률분포에 대한 에너지 함수이다. DEM이 GBRBM의 심층 구조 형태이듯이, 제안된 EBLVM은 KRBM^[17]에 대한 심층 구조를 갖는 일반적인 형태로 생각할 수 있다.

제안된 EBLVM의 파라미터를 학습하기 위해서는 식(1)의 EBM과 같이 로그 유사도를 근사적으로 최대화하는 CD-1 알고리즘^[10-12]을 다음과 같이 사용한다.

$$\begin{aligned} \nabla_{\theta} L(\theta) = & -\mathbf{E}_{\mathbf{x} \sim p_{data}(\mathbf{x}), \mathbf{h} \sim p_{\theta}(\mathbf{h} | \mathbf{x})} [\nabla_{\theta} E_{\theta}(\mathbf{x}, \mathbf{h})] \\ & + \mathbf{E}_{\mathbf{x}' \sim p_{\theta}(\mathbf{x}'), \mathbf{h}' \sim p_{\theta}(\mathbf{h}' | \mathbf{x}')} [\nabla_{\theta} E_{\theta}(\mathbf{x}', \mathbf{h}')] \end{aligned} \quad (9)$$

여기서 위 식의 오른쪽 첫째 항은 학습 데이터로부터 쉽게 구할 수 있는 항이며, 두 번째 항의 \mathbf{x}' 와 \mathbf{h}' 는 모델로부터 생성된 가시변수와 은닉변수의 샘플이다. 그리고 식(5)으로부터 $\nabla_{\theta} E_{\theta}(\mathbf{x}, \mathbf{h}) = -\sum_i h_i \nabla_{\theta} \mathbf{f}_{\theta}(\mathbf{x})_i$

로 주어지며, 여기서 h_i 와 $\mathbf{f}_{\theta}(\mathbf{x})_i$ 은 각각의 i 번째 성분을 나타낸다. 식(9)의 두 번째 항을 계산하기 위해서는 모델로부터 생성된 새로운 데이터 \mathbf{x}' 와 \mathbf{h}' 가 필요하다. 최근에 가장 많이 사용되는 샘플링 기법은 앞에서 언급된 SGLD이다. 그러나 여기서는 \mathbf{x}' 와 \mathbf{h}' 를 반복적인 과정을 통해 샘플링하는 깃스샘플링 과정을 수행한다. 먼저 \mathbf{h}' 는 식(6)으로부터 간단하게 샘플링 된다. \mathbf{x}' 을 생성하기 위해서는 식(7)의 로그 조건부 확률을 최대화하도록, $\mathbf{x}' = \arg\max_{\mathbf{x}} \ln p(\mathbf{x} | \mathbf{h})$ 로 샘플링 한다^[17]. 이를 이용하여 본 논문에서는 [17]와 같은 방식으로 아래와 같은 간단한 샘플링 기법을 사용한다.

$$\mathbf{x}' = \sum_i h_i \nabla_{\theta} \mathbf{f}_{\theta}(\mathbf{x})_i \quad (10)$$

여기서 \mathbf{x}' 는 \mathbf{h} 가 주어진 경우 생성된 샘플이며, $\nabla_{\theta} \mathbf{f}_{\theta}(\mathbf{x})_i$ 는 심층신경망 $\mathbf{f}_{\theta}(\mathbf{x})$ 의 i 번째 은닉변수 출력에 대한 입력벡터의 경사도를 나타낸다. 위와 같은 데이터의 샘플링은 실험 결과 매우 효과적임을 알 수 있다.

새로운 EBLVM을 학습할 때는 DEM와 같이 layer-wise 학습과 연합학습(또는 fine-tuning이라 함) 방식으로 번갈아 진행된다^[6]. 예를 들면 2개의 은닉층을 갖는 경우, 우선 첫 번째 층을 학습한 후, 첫 번째 층의 출력을 받아 2번째 층을 layer-wise 방식으로 학습한다. 그리고 나서 2개의 층 모두를 연합방식으로 학습하여 모든 층의 파라미터를 갱신한다. 이런 방식의 학습은 DBM과 달리 중간계층의 은닉유닛의 샘플링이 필요가 없게 된다. 그리고 이러한 학습 방식은 모델의 수렴을 더욱 좋게 할 뿐만 아니라 전체적으로 최적의 파라미터를 얻게 되어 모든 층에서 더 효과적인 표현을 추출할 수 있다. 새로운 EBLVM을 위한 학습 알고리즘을 요약하면 다음과 같다.

제안된 EBLVM 학습 알고리즘(CD-1 기반)
 입력: 학습하기 위한 batch data S, 학습률 η
 출력: 파라미터의 경사값

1. θ 값 초기화
2. for all the $\mathbf{x} \in S$ do
3. $\mathbf{f}_\theta(\mathbf{x})$ 계산
4. 식(6)을 이용해 \mathbf{h} sampling
5. 식(10)을 이용해 \mathbf{x}' sampling
6. 식(6)을 이용해 \mathbf{h}' sampling
7. 식(9)이용, $\theta \leftarrow \theta + \eta \nabla_{\theta} L(\theta)$ 로 파라미터 갱신

III. 실험

새로운 EBLVM의 성능을 평가하기 위해 두 가지 데이터 셋, Fashion MNIST^[19]와 CIFAR10^[20]에 대한 비지도 특징학습을 통한 이미지 인식 실험을 수행하였다.

3.1 Fashion MNIST

Fashion MNIST 데이터 세트는 총 70,000개로 구성된 28x28 크기의 gray 이미지이다. 총 10개 클래스가 있고 60,000장의 훈련 이미지와 10,000장의 테스트 이미지로 구성된다. 제안된 EBLVM 기법의 성능을 비교하기 위해 stacked KRBM, DBN, DBM 그리고 DEM에 대해 동일한 훈련과 인식실험을 수행하였다. Fashion MNIST의 훈련 데이터를 이용하여 각각의 EBLVM을 훈련시킨 후, 훈련과 테스트 데이터를 사용하여 각 EBLVM의 은닉층에서 특징들을 추출하고, 이들을 softmax 분류기의 훈련 및 인식실험에 이용하였다. 각각의 모델은 최대 4층의 은닉층 구조를 가지며, 각 층은 L1, L2, L3, L4로 표시된다. CD-1 알고리즘을 사용하여 새로운 EBLVM과 stacked KRBM 모델을 layer-wise 방식으로 훈련시켰을 때 각 계층의 학습률은 다음과 같다: L1은 1e-4, 나머지 층: 1e-5이다. 새로운 EBLVM 모델을 연합훈련 방식으로 훈련시켰을 때 학습률은 (L1, L2): 1e-5, (L1, L2, L3): 1e-6, (L1, L2, L3, L4): 1e-7이다. DBN와 DBM 모델의 학습률은 L1: 1e-4, L2: 1e-5, 나머지 층: 1e-6이다. 연합 훈련 방식으로 DEM을 훈련시켰을 때, 각 계층의 학습률은 (L1, L2): 1e-5, (L1, L2, L3): 1e-6, (L1, L2, L3, L4): 1e-7이다. 각 모델의 구현은 TensorFlow로 작성하였고, Adam optimizer을 사용하

였다. 모든 모델에 대해 batch size는 100이고, 총 500 epoch 훈련되며, stacked KRBM, DBN과 DBM에 대한 20 epoch 단위로 모델의 파라미터가 저장하고, 새로운 EBLVM과 DEM에 대해 연합훈련은 10 epoch을 수행하여 마지막에 파라미터가 저장했다. 입력층의 크기는 28x28=784이며, 다양한 은닉층 크기를 갖는 심층신경망 사용이 가능하지만 본 실험에서는 모든 모델에 대해 (1024-512-512-256) 구조의 은닉층을 사용하였다. 각 은닉층의 activation 함수는 stacked KRBM과 제안된 모델에 대해서는 ReLU(Retified Linear Unit)을 사용하였고, 나머지 모델은 sigmoid 함수를 사용하였다. 훈련이 끝난 후 stacked KRBM, DBN과 DBM에 대해 각 은닉층에서 특징을 추출하고, 이를 연결하여, 이를 softmax 분류기로 보내 훈련 및 인식을 수행하였다. 새로운 EBLVM과 DEM에 대해 연합 훈련된 파라미터를 이용하여 똑같은 실험을 수행하였다. 그림 2에 각 모델에 대해 특징학습을 수행한 후 인식성능의 결과를 나타내었다. 한 층에 대해서는 DBN, DBM 그리고 DEM의 성능은 같고, 또한 KRBM과 제안된 EBLVM의 성능은 같다. DBN과 DEM은 은닉층이 증가함에 따라 성능도 향상되었고, 3-4층에서 성능이 비슷하였다. 그러나 DBM은 2층의 경우에는 성능이 향상되었으나, 3-4층의 경우 오히려 성능이 저하되었다. 이는 층이 많아짐에 따라 overfitting 현상이 DBM에서 발생한 것으로 생각된다. Stacked KRBM과 제안된 EBLVM은 다른 모델에 비해 더 나은 성능을 나타내었으며, 특히 제안된 모델이 미리 학습된 stacked KRBM을 fine-tuning 형식으로 모든 층을 동시에 훈련함으로써 stacked KRBM보다 조금 더 우수한 인식성능을 나타내었다.

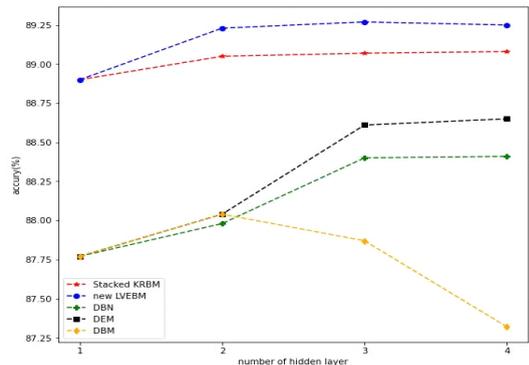


그림 2. Fashion MNIST에서 은닉층 수에 따른 테스트 인식 정확도.
 Fig. 2. Test recognition accuracy as the number of hidden layers in Fashion MNIST.

새로운 EBLVM은 세 개의 은닉층을 가지고 있을 때 성능이 가장 좋았고 4층으로 올라갔을 때 overfitting 때문에 약간의 성능이 저하되었다. 전체적으로 제안된 모델이 2층 이상의 은닉층을 가진 구조에서 기존 기법에 비해 가장 뛰어난 인식성능을 나타내었다. 따라서 새로운 EBLVM은 Fashion MNIST를 사용한 특징 학습에 있어서 매우 효과적임을 알 수 있었다. 이러한 EBLVM의 특징은 여러층을 쌓아 연합하여 훈련함으로써 더 많은 유용한 특징들을 추출할 수 있는 반면, 모델의 파라미터 수가 증가하여 계산량과 메모리 상에 대한 복잡도가 증가하는 단점을 갖는다.

3.2 CIFAR10

CIFAR10 데이터 셋은 60,000개의 32x32 크기의 컬러 이미지를 포함하고 있으며, 10개의 클래스 (airplane, bird, automobile, cat, deer, dog, horse, frog, ship, truck)로 구성되어 있다. 훈련 이미지는 50,000장, 테스트 이미지는 10,000장이다. 모든 모델에 대해 훈련할 때, 은닉층의 개수, 학습률, 가중치 초기화, batch size, epoch 설정은 Fashion MNIST와 거의 동일하게 사용하였다. 단 DEM에 대한 연합훈련을 진행할 때 학습률은 (L1, L2, L3): 0.5e-8이고 (L1, L2, L3, L4): 0.5e-9을 사용하였다. 입력층의 크기는 32x32x2=3,072이며, 4개의 은닉층에 있는 뉴런 수는 (5000-1000-512-256)으로 설정하였다. 새로운 EBLVM과 stacked KRBM을 훈련하기 전에, 데이터는 영 평균을 갖도록 처리하였으며, 나머지 모델에 대해 데이터를 정규화하여, 모델에 입력된 데이터가 0 평균과 단위 분산을 갖도록 하였다. 훈련이 끝난 후 앞에서 언급한 실험과 마찬가지로 분류실험을 진행하였다. 그림 3은 CIFAR10에 대해 각 모델에 대한 결과이다. 전체적으로 Fashion MNIST와 비슷한 결과를 나타내었다. 다만 DBN, DBM 그리고 DEM에 대해서 은닉층의 수가 증가 함에 따라 성능향상이 그리 크지 않고 비슷한 성능을 나타내었다. 그러나 stacked KRBM과 제안된 EBLVM은 다른 기법에 비해 많은 인식성능 향상을 보였다. 새로운 EBLVM과 stacked KRBM과 비교하였을 때, 새로운 EBLVM이 stacked KRBM에 비해 약간의 성능 향상을 확인할 수 있었다. 마찬가지로 새로운 EBLVM이 stacked KRBM을 모든 층의 파라미터를 동시에 fine-tuning한 결과로 성능 향상에 기여한 것으로 판단된다. 새로운 EBLVM은 3 개의 은닉층으로 구성하였을 때 성능이 가장 좋지만 4 층의 구조에서는 약간의 성능이 하락하였다. 이는 은닉층이 많아 질수록 파라미터수가 증가하고 따라서

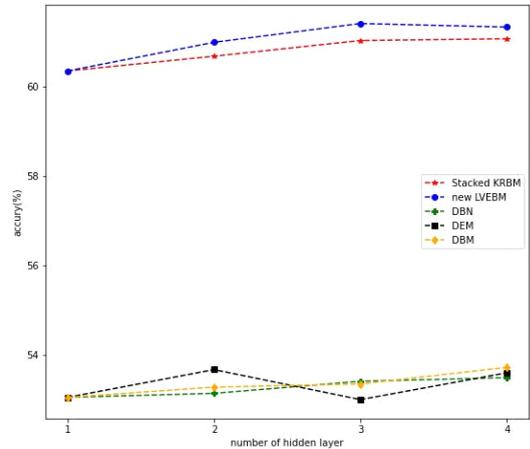


그림 3. CIFAR10에서 은닉층 수에 따른 테스트 인식 정확도. Fig. 3. Test recognition accuracy as the number of hidden layers in CIFAR10.

전체적으로 overfitting이 발생한 것으로 생각된다. CIFAR10의 인식실험 결과 최고 정확도는 60% 부근에서 나타났다. CNN을 이용한 지도학습의 경우 인식률이 90% 이상에 도달할 수 있다. 그러나 본 논문의 모델은 feed-forward 신경망 구조를 가지므로 이미지를 벡터형태로 바꾸어 신경망에서 사용함으로써 성능저하를 유발하고 있다. 비록 구조적 한계로 인해 인식실험에서 높은 성능을 달성하지 못했지만, 특징학습에서 기존 모델에 비해 더 높은 성능을 나타내고 있다. 전체적인 결과를 살펴볼 때 새로운 EBLVM이 CIFAR10 데이터 셋에서도 특징학습을 통한 인식실험에서 다른 모델에 비해 효과적임을 알 수 있었다.

IV. 결 론

본 논문에서는 비지도 특징학습을 위한 새로운 은닉변수 에너지기반 모델(EBLVM)을 제시하였다. 제안한 EBLVM은 결합 확률밀도함수를 정의하기 위해 심층 feedforward 신경망을 이용해 변환된 가시변수와 은닉변수에 대해 새로운 에너지 함수를 모델링 하였다. 이러한 EBLVM은 경사기반 contrastive divergence 알고리즘을 사용하여 파라미터를 학습하였다. 새로운 EBLVM은 심층 구조가 갖고 여러 층을 연합하여 훈련하기 때문에 각 층에서 더 유용한 특징들을 추출할 수 있었다. Fashion MNIST와 CIFAR10 데이터를 사용한 비지도 특징학습을 통한 이미지 인식 실험에서 제안된 기법은 기존의 stacked RBM, DBN, DBM 그리고 DEM보다 더 우수한 인식결과를

나타내었다.

앞으로의 연구 방향으로는 제안된 구조하에서 convolutional 형태의 EBLVM 연구가 필요하며, 이를 음성, 텍스트 등 다양한 형태의 데이터에 적용이 필요하다. 그리고 생성모델에 이 구조를 적용하는 연구를 수행함으로써 다양한 데이터의 생성에 이용하고, 최근 모델들과 비교할 예정이다.

References

- [1] R. Hankins, Y. Peng, & H. Yin, "SOMNet: unsupervised feature learning networks for image classification," in *IEEE, IJCNN*, pp. 1-8, 2018. (<https://doi.org/10.1109/IJCNN.2018.8489404>)
- [2] J. Schlüter, & R. Sonnleitner, "Unsupervised feature learning for speech and music detection in radio broadcasts," in *Proc. 15th ICDAE*, September, 2012.
- [3] V. Christlein, M. Gropp, S. Fiel, & A. Maier, "Unsupervised feature learning for writer identification and writer retrieval," *IEEE, 14th ICDAR*, Vol. 1, pp. 991-997, 2017. (<https://doi.org/10.48550/arXiv.1705.09369>)
- [4] Y. LeCun, S. Chopra, R. Hadsell, M. Ranzato, & F. Huang. "A tutorial on energy-based learning," *Predicting Structured Data*, 1.0, 2006.
- [5] Y. Song, & D. P. Kingma, "How to train your energy-based models," *arXiv preprint arXiv:2101.03288*, 2021. (<https://doi.org/10.48550/arXiv.2101.03288>)
- [6] J. Ngiam, Z. Chen, P. W. Koh, & A. Y. Ng, "Learning deep energy models," *Proc. 28th ICML*, pp. 1105-1112, 2011.
- [7] F. K. Gustafsson, M. Danelljan, R. Timofte, & T. B. Schön, "How to train your energy-based model for regression," *arXiv preprint arXiv:2005.01698*, 2020. (<https://doi.org/10.48550/arXiv.2005.01698>)
- [8] M. M. Al Rahhal, Y. Bazi, R. Al-Dayil, B. M. Alwadei, N. Ammour, & N. Alajlan, "Energy-based learning for open-set classification in remote sensing imagery," *IJRS*, 43.15-16, pp. 6027-6037, 2022. (<https://doi.org/10.1080/01431161.2022.2044539>)
- [9] Y. Du, & I. Mordatch, "Implicit generation and modeling with energy based models," *Advances in Neural Information Processing Systems* 32, 2019. (<https://doi.org/10.48550/arXiv.1903.08689>)
- [10] H. Wu, B. Esmacili, M. Wick, J. B. Tristan, & J. W. Van De Meent, "Conjugate energy-based models," *ICML*, pp. 11228-11239, 2021. (<https://doi.org/10.48550/arXiv.2106.13798>)
- [11] G. E. Hinton, "A practical guide to training restricted Boltzmann machines," *Tech. Rep. UTM TR2010-003*, University of Toronto, 2010. (https://doi.org/10.1007/978-3-642-35289-8_32)
- [12] Fischer, Asja, and Christian Igel. "Training restricted Boltzmann machines: an introduction." *Pattern Recognition*, pp. 25-39, 2014. (<https://doi.org/10.1016/j.patcog.2013.05.025>)
- [13] K. Cho, A. Ilin, & T. Raiko, "Improved learning of Gaussian-Bernoulli restricted Boltzmann machines," in *proc. 21th ICANN*, pp. 10-17. Springer Berlin Heidelberg, 2011. (https://doi.org/10.1007/978-3-642-21735-7_2)
- [14] V. Nair, & G. E. Hinton, "Rectified linear units improve restricted Boltzmann machines," *Proc. 27th ICML*, pp. 807-814, 2010.
- [15] G. E. Hinton, S. Osindero, & Y. W. Teh, "A fast learning algorithm for deep belief nets," *Neural computation*, 18(7), pp. 1527-1554, 2006. (<https://doi.org/10.1162/neco.2006.18.7.1527>)
- [16] R. Salakhutdinov, & H. Larochelle, "Efficient learning of deep Boltzmann machines," *Proc. thirteenth ICARS, PMLR* 15, pp. 693-700, 2010.
- [17] D. Kim & J. W. Shin, "Kernel-based restricted Boltzmann machine for unsupervised feature learning," *J. KICS*, Vol. 44, No. 09, pp. 1633-1640, 2019. (<https://doi.org/10.7840/kics.2019.44.9.1633>)
- [18] D. Kim, "Generalized kernel restricted Boltzmann machine" *J. KICS*, Vol. 45, No.

05, pp. 1633 - 1640, 2020.

(<https://doi.org/10.7840/kics.2020.45.5.783>)

- [19] H. Xiao, K. Rasul, & R. Vollgraf, "Fashion-mnist: a novel image dataset for benchmarking machine learning algorithms," *arXiv preprint arXiv:1708.07747*, 2017 (<https://doi.org/10.48550/arXiv.1708.07747>)
- [20] A. Krizhevsky, & G. Hinton, "Learning multiple layers of features from tiny images," 2009.

곽 봉 (Guo Peng)



2014년 9월 : 전남대학교 전자공학과 학사

2018년 9월 : 전남대학교 전자공학과 석사

2021년 3월~현재 : 전남대학교 전자공학과 박사과정

<관심분야> 영상처리, 기계학습, 딥러닝

[ORCID:0009-0009-0123-9957]

김 등 국 (Dong Kook Kim)



1989년 2월 : 전남대학교 전자공학과 학사

1991년 2월 : 포항공과대학 전자전기공학과 석사

2003년 2월 : 서울대학교 전기컴퓨터공학부 박사

1991년 2월~1999년 2월 : 삼성전자 전문연구원

2003년 4월~2004년 2월 : 한국전자통신연구원 선임연구원

2004년 2월~현재 : 전남대학교 전자공학과 교수

<관심분야> 딥러닝, 기계학습, 인공지능신호처리

[ORCID:0000-0001-9316-7069]