https://doi.org/10.7840/kics.2025.50.10.1497

이상탐지를 위한 결합 에너지 기반 모델의 잡음 대비 추정

김 동 국◆

Noise Contrastive Estimation of Coupled Energy-Based Model for Anomaly Detection

Dong Kook Kim◆

요 약

본 논문에서는 이상탐지를 위한 새로운 결합에너지기반모델(CEBM)과 이를 학습하기 위한 잡음대비추정(NCE) 기법을 제시한다. CEBM의 구조는 비정규화된 두 개의 확률분포의 곱으로 구성되며, 각각의 분포는 심층신경망을 갖는 에너지 함수에 의해 정의된다. CEBM의 하나의 분포를 잡음분포로 사용하여 NCE의 목적함수를 유도하고 경사하강법에 따른 파라미터의 갱신법을 제시한다. 제안된 기법의 이상탐지 성능을 평가하기 위해 ECG, UNSW 그리고 MNIST/Fashion-MNIST를 이용한 실험을 수행한다. 실험 결과로 제안된 NCE로 학습된 CEBM이 모든 데이터 세트에서 기존의 EBM보다 더 높은 F1-score를 보여준다.

Key Words : energy-based model, noise contrastive estimation, anomaly detection, deep neural networks

ABSTRACT

In this paper, a new coupled energy-based model (CEBM) for anomaly detection and noise contrast estimation (NCE) technique for learning it are presented. The structure of CEBM consists of the product of two unnormalized probability distributions, each of which is defined by an energy function with a deep neural network. We derive the objective function of NCE by using one distribution of CEBM as a noise distribution and present a parameter update method using the gradient descent method. An experiment using ECG, UNSW and MNIST/Fashion-MNIST is conducted to evaluate the anomaly detection performance of the proposed technique. As a result of the experiment, the CEBM learned with the proposed NCE shows higher F1-score than the existing EBMs in all data sets.

Ⅰ. 서 론

에너지 기반 모델 (Energy-based Models, EBMs)[1-4]은 기계학습 비지도 학습(unsupervised learning) 분야에서 널리 사용되는 하나의 모델로 최근 관련 연구가 활발히 진행되고 있다. EBM은 심층신경망을 사용하는 에너지 함수(energy function)를 통해 입력 데이터의 분포를 모델링하는 확률적인 기법이다. 에너지 함수로 복잡한 심층신경망을 사용하기 때문에 EBM는 확률분포에 대한 비정규화된(unnormalized) 파티션 함수(partition function)을 갖는 것이 특징이다[2]. 이러한 EBM의 구조로 인해 다양한 데이터의 분포를 효과적으

◆ First Author : Chonnam National University, School of Electronics and Computer Engineering, dkim@jnu.ac.kr, 종신회원
논문번호 : 202504-090-A-RE, Received April 18, 2025; Revised May 22, 2025; Accepted May 31, 2025

로 모델링할 수 있어 특징학습, 이미지 생성, 밀도추정 그리고 이상탐지 (anomaly detection) 등과 같은 다양한 분야에 응용되고 있다[5-7].

EBM은 심층신경망을 기반으로 고차원의 복잡한 데이터의 분포를 정밀하게 표현할 수 있어 이상탐지 분야에서 널리 사용되고 있다[5]-[8]. 특히 정상 데이터만을 사용하여 분포를 모델링하고 이 분포를 이용하여 정상/비정상을 탐지하는 준지도(semi-supervised) 이상탐지에 EBM 구조가 매우 적합하다[8]. 정상 데이터의 수집은 상대적으로 쉽지만, 이상 데이터의 획득이 어렵고 데이터 labeling의 큰 노력이 필요하므로 준지도 이상탐지에 autoencoders(AEs)[9], generative adversarial networks (GANs)[10]과 함께 EBM이 널리 쓰이고 있다.

EBM는 비정규화된 분포 형태로 인하여 데이터에 대한 정확한 우도비(likelihood) 계산이나 데이터 생성 과정에 많은 계산량이 필요하여 학습이 매우 어려운 것으로 알려졌다[2]. EBM의 파라미터 학습기법에는 Markov chain Monte Carlo(MCMC)[11]에 기반한 최대우도비(Maximum－Likelihood, ML), 스코어 매칭(Score Matching, SM) 그리고 잡음대비추정(Noise Contrastive Estimation, NCE) 등이 있다[2]. ML 학습은 파티션 함수의 문제점을 극복하기 위해 MCMC를 사용하여 데이터를 생성하고 이를 학습에 이용하는 기법이다. 그러나 이는 데이터 생성 과정에 매우 많은 계산량이 요구되는 단점이 있다. 이러한 단점을 극복하기 위해 학습 데이터를 초기값으로 하여 짧은 MCMC를 수행하는 contrastive divergence(CD)[4] 기법이 사용되고 있다.

SM[2,12-14] 기법은 파티션 함수를 계산하지 않고 EBM을 학습할 수 있는 기법이다. SM의 목적함수는 데이터 분포와 모델 분포 사이에 Fisher divergence라 불리는 두 분포의 로그 미분값 차이의 유클리드 거리(Euclidean distance)에 의해 정의된다. 이 기법의 장점은 MCMC와 같은 표본화 과정이 필요 없다는 것이다. 그러나 목적함수 내의 2차 도함수 때문에 많은 계산이 요구되는 단점이 있다.

NCE[2,15-18]은 알려진 잡음분포와 EBM을 대비하여 학습하는 기법이다. NCE의 특징 중의 하나는 파티션 함수를 파라미터로 취급하여 학습 중에 추정할 수 있으며, 이에 따라 데이터의 우도비 계산이 가능하다는 것이다[15]. 그러나 NCE 단점은 잡음분포의 선택에 따라 성능이 매우 큰 영향을 받는 것이다. 잡음분포가 데이터 분포와 완전히 다른 경우는 학습이 잘 안되지만, 근사적으로 가까울 때 더 좋은 성능을 나타낸다[14]. 잡음분포를 데이터 분포와 가깝게 학습하는 기법들이 제안되었지만. 사용되는 분포의 형태가 매우 제한적이라는 단점을 갖는다[16-18].

본 논문에서는 두 가지 확률분포를 결합한 새로운 EBM 구조와 이를 NCE 기법으로 학습하는 알고리즘을 제안한다. 두 가지 확률분포는 비정규화된 EBM 구조를 가지며, 하나는 변별기(discriminator)로, 다른 하나는 NCE의 잡음분포의 역할을 수행한다. 각각 EBM의 에너지 함수는 여러 형태의 심층신경망에 의해 정의된다. 제안된 구조의 NCE 학습을 위해 GAN과 비슷한 최소최대(minmax) 손실함수를 유도하고 gradient descent(GD) 알고리즘에 근거한 파라미터를 갱신법을 제시한다. 제안된 학습기법의 성능을 평가하기 위해 준지도 이상탐지 실험을 수행한다. 이상탐지 실험을 위해 ECG[19], UNSW[20] 그리고 MNIST[21]/Fashion-MNIST[22] 데이터 셋을 사용한다. 실험 결과 제안된 기법은 모든 데이터 세트에서 AE, ML와 SM 기반 EBM보다 더 높은 F1-score 성능을 나타내었다. 또한 다른 기법과의 성능 비교를 통해 제안된 CEBM과 NCE 학습기법이 이상탐지에 매우 효과적임을 나타낸다.

본 논문의 본론 Ⅱ장에서는 EBM과 NCE 기법을 소개하고, 결합 EBM과 이를 NCE로 학습하는 알고리즘을 제시한다. Ⅲ장에서는 실험 및 결과를 나타내고, Ⅳ에서는 결론을 맺는다.

Ⅱ. 본 론

2.1 EBM과 학습기법

이 장에서는 EBM과 이를 학습하기 위한 NCE 기법을 간단히 소개한다. 확률변수 $\boldsymbol{x}$을 입력신호라 하고, $\boldsymbol{x}$의 모델 분포를 파라미터 θ을 갖는 $p_\theta(\boldsymbol{x})$라 하자. EBM은 $p_\theta(\boldsymbol{x})$를 다음과 같은 Boltzmann 분포로 정의한다[1,2].

$$p_\theta(\boldsymbol{x}) = \frac{e^{-E_\theta(\boldsymbol{x})}}{Z_\theta} \tag{1}$$

여기서 $E_\theta(\boldsymbol{x})$는 에너지 함수이며, 파라미터 θ를 갖는 심층신경망에 의해 정의된다. $Z_\theta = \int \exp\{-E_\theta(\boldsymbol{x})\}d\boldsymbol{x}$는 θ의 함수로 정규화 상수 또는 파티션 함수라 한다. EBM의 $p_\theta(\boldsymbol{x})$가 주어진 경우 이로부터 샘플들은 MCMC 기법을 사용하여 생성한다. EBM을 위한 가장 일반적인 MCMC 알고리즘은 stochastic gradient Langevin dynamics (SGLD) 기법이며, 아래와 같이 반복적인 업데이트 식을 사용된다[11].

$$\boldsymbol{x}_t = \boldsymbol{x}_{t-1} - \alpha_t \nabla_{\boldsymbol{x}} E_\theta(\boldsymbol{x_{t-1}}) + \sqrt{2\alpha_t}\epsilon_t\,,\ \epsilon_t \sim N(0,I) \tag{2}$$

여기서 $\alpha_t > 0$은 step-size이며, $N(0,I)$ 은 평균 0, 공분산이 단위행렬 I인 다변수 정규분포이다. 샘플은 초기값 $\boldsymbol{x}_0$에서 시작하여 단계별로 반복적인 가산적인 가우시안 잡음을 더한 후 에너지 함수의 경사도 값을 통해 갱신된다. 다양한 가정하에서 매우 작은 α_t와 큰 시간 T에 대해 $\boldsymbol{x}_T$의 분포가 $p_\theta(\boldsymbol{x})$에 수렴하는 것으로 알려졌다[11].

EBM을 학습하기 위한 대표적인 방법으로 MCMC 기반 ML, SM 그리고 NCE 기법이 있다[2]. ML와 SM 기법이 가장 많이 사용되지만, 최근에 NCE 기법이 그 장점 때문에 주목을 받고 있다[2,15-18]. NCE의 핵심 아이디어는 EBM의 학습이 이미 알려진 잡음분포와 대비하여 이루어지는 것이다[2]. $p_{data}(\boldsymbol{x})$를 데이터 분포, $p_\theta(\boldsymbol{x})$를 EBM 그리고 $p_n(\boldsymbol{x})$를 잡음분포라 하자. 이때 잡음분포는 $N(0,I)$ 와 같이 간단하며, 효율적으로 분포 값을 계산할 수 있고 또한 샘플링할 수 있는 분포를 사용한다. NCE 학습은 관측 데이터와 잡음 분포로부터 생성된 데이터를 서로 대비시키는 이진 분류기(logistic regression)를 사용하여 두 데이터가 최대 분류가 되도록 파라미터를 추정한다. 관측 데이터와 잡음 데이터에 대한 선행 확률이 같다고 가정하면, NCE의 목적함수는 아래와 같이 주어진다[2,15].

$$\begin{aligned} &V(\theta) \\ &= E_{p_{data}(\boldsymbol{x})}\left[\log\frac{p_\theta(\boldsymbol{x})}{p_\theta(\boldsymbol{x})+p_n(\boldsymbol{x})}\right] + E_{p_n(\boldsymbol{x})}\left[\log\frac{p_n(\boldsymbol{x})}{p_\theta(\boldsymbol{x})+p_n(\boldsymbol{x})}\right] \end{aligned} \tag{3}$$

다른 학습기법에 비해 NCE의 특징은 정규화 상수 Z_θ를 파라미터로 학습이 가능하다는 것이다. 위의 목적함수를 최대화하도록 파라미터를 학습하면 $p_\theta(\boldsymbol{x})$가 데이터 분포 $p_{data}(\boldsymbol{x})$를 근사화하는 것으로 알려져 있다. NCE의 단점은 성능이 잡음분포 $p_n(\boldsymbol{x})$의 선택에 따라 매우 큰 영향을 받는다는 것이다. $p_n(\boldsymbol{x})$은 다음 3가지 조건을 만족하도록 선택된다[15][18]. (i) 정규화된 분포함수로 분포 값의 계산이 가능해야 하며, (ii) 이 분포로부터 샘플링이 쉬어야 하며, (iii) 데이터 분포와 비슷한 형태를 가져야 한다는 것이다. (i)와 (ii)을 조건을 만족해야 목적함수 식(3)의 계산이 가능하며, (iii)의 조건을 만족해야 좋은 성능을 나타내게 된다. 특히 이미지와 같은 고차원의 데이터에 대해 단순한 가우시안 분포를 잡음분포로 사용하는 경우 학습이 어려운 것으로 알려졌다[15]. 따라서 NCE 기법으로 효과적으로 학습하기 위해 EBM의 구조와 잡음분포 추정에 관한 연구가 요구되고 있다.

2.2 결합 EBM과 NCE 학습

이 단원에서는 EBM의 새로운 형태인 결합 EBM과 이를 학습하기 위한 NCE 기법을 제안한다.

새롭게 제시하는 결합 EBM의 구조는 비정규화된 두 개의 분포, $p_\theta(\boldsymbol{x})$와 $q_\phi(\boldsymbol{x})$의 곱으로 정의되는 확률모델이다.

$$h_{\theta,\phi}(\boldsymbol{x}) = \frac{p_\theta(\boldsymbol{x})q_\phi(\boldsymbol{x})}{Z_{\theta,\phi}} \tag{4}$$

여기서 $p_\theta(\boldsymbol{x})$와 $q_\phi(\boldsymbol{x})$는 각각 파라미터 θ와 ϕ을 갖는 비정규화된 EBM 구조, 즉 $p_\theta(\boldsymbol{x}) \propto e^{-D_\theta(\boldsymbol{x})}$와 $q_\phi(\boldsymbol{x}) \propto e^{-G_\phi(\boldsymbol{x})}$을 갖는다고 가정한다. 정규화 상수 $Z_{\theta,\phi}$는 확률분포의 정규화 조건을 만족하기 위해 $Z_{\theta,\phi} = \int p_\theta(\boldsymbol{x})q_\phi(\boldsymbol{x})d\boldsymbol{x}$이며, θ,ϕ의 함수이다. CEBM의 분포 $h_{\theta,\phi}(\boldsymbol{x})$는 $p_\theta(\boldsymbol{x})$와 $q_\phi(\boldsymbol{x})$의 결합 형태로, 각각의 에너지 함수 $D_\theta(\boldsymbol{x})$와 $G_\phi(\boldsymbol{x})$가 심층신경망으로 구성된다. 따라서 이를 결합 에너지기반 모델(Coupled EBM, CEBM)이라 부른다. 이러한 CEBM의 특징은 서로 다른 역할을 수행하는 분포들을 결합함으로써 데이터를 효과적으로 모델링할 수 있고, NCE 학습과정이 더 단순하게 수행되는 장점이 있다.

여기서 CEBM 구조에 기반하여 $h_{\theta,\phi}(\boldsymbol{x})$가 데이터분포 $p_{data}(\boldsymbol{x})$에 근사화되도록 파라미터 θ,ϕ를 추정하는 NCE 기법을 제시한다. 이를 위해 본 논문에서는 CEBM의 일부인 $q_\phi(\boldsymbol{x})$를 잡음분포로 사용한다. 이에 따라 모델과 잡음분포의 파라미터 θ와 ϕ를 식(3)에 근거하여 반복적인 방법에 따라 학습한다. 이때 NCE는 GAN과 비슷한 최소최대 손실함수, $\min_{\phi}\max_{\theta} V(\theta,\phi)$을 사용하고, 다음 식으로 유도된다[17][18].

$$\begin{aligned} &V(\theta,\phi) \\ &= E_{p_{data}(\boldsymbol{x})}\left[\log\frac{h_{\theta,\phi}(\boldsymbol{x})}{h_{\theta,\phi}(\boldsymbol{x})+q_\phi(\boldsymbol{x})}\right] + E_{q_\phi(\boldsymbol{x})}\left[\log\frac{q_\phi(\boldsymbol{x})}{h_{\theta,\phi}(\boldsymbol{x})+q_\phi(\boldsymbol{x})}\right] \\ &= E_{p_{data}(\boldsymbol{x})}\left[\log\frac{p_\theta(\boldsymbol{x})}{p_\theta(\boldsymbol{x})+1}\right] + E_{q_\phi(\boldsymbol{x})}\left[\log\frac{1}{p_\theta(\boldsymbol{x})+1}\right] \\ &= E_{p_{data}(\boldsymbol{x})}\left[\log\sigma(-D_\theta(x))\right] + E_{q_\phi(\boldsymbol{x})}\left[\log(1-\sigma(-D_\theta(x)))\right] \end{aligned} \tag{5}$$

여기서 $\sigma(x)=1/(1+e^{-x})$는 sigmoid 함수이다. 위의 첫 번째 식은 식(3)에서 $p_\theta(\boldsymbol{x})$대신에 $h_{\theta,\phi}(\boldsymbol{x})$를, $p_n(\boldsymbol{x})$대신 $q_\phi(\boldsymbol{x})$를 대입함으로 얻을 수 있다. 두 번째 식에서는 $q_\phi(\boldsymbol{x})$가 상쇄되며, 마지막 식은 간단한 sigmoid 함수식에 의해 정의된다. 식(5)은 변별기와 생성기(generator)을 갖는 GAN 손실함수와 매우 비슷하다[10]. $D_\theta(\boldsymbol{x})$는 학습 데이터 샘플과 $q_\phi(\boldsymbol{x})$로부터 생성되는 샘플을 판별하는 변별기 역할을 수행하며, $G_\phi(\boldsymbol{x})$는 데이터 분포를 근사적으로 추정하는 잡음분포의 역할을 수행한다. 따라서 CEBM은 두 가지 역할을 수행하는 모델들이 하나로 결합된 구조를 갖는 것이 특징이다. 또한 손실함수에서 $q_\phi(\boldsymbol{x})$의 분포 값 계산을 필요하지 않아 잡음분포의 다양한 선택이 가능하다는 장점이 있다.

그러나 $q_\phi(\boldsymbol{x})$을 잡음분포로 사용함으로써 샘플링이 더 복잡해지는 단점이 있다. 샘플링 과정은 $q_\phi(\boldsymbol{x})$의 에너지 함수 $G_\phi(\boldsymbol{x})$ 구조와 MCMC 기법에 의존한다. 본 논문에서는 샘플링을 위해 식(2)의 SGLD와 더블어 아래와 같은 에너지 함수 $G_\phi(\boldsymbol{x})$를 사용한다[6].

$$q_\phi(\boldsymbol{x})=e^{-G_\phi(\boldsymbol{x})},\ G_\phi(\boldsymbol{x})=\frac{1}{2}\left(\boldsymbol{x}^T\boldsymbol{x}-\boldsymbol{f}_\phi^T(\boldsymbol{x})\boldsymbol{f}_\phi(\boldsymbol{x})\right) \tag{6}$$

여기서 $\boldsymbol{f}_\phi(\boldsymbol{x})$은 여러 개의 출력 노드를 갖는 심층신경망을 나타낸다. 식(2)의 SGLD를 이용한 샘플링을 수행하기 위해 $\boldsymbol{x}$에 대한 $G_\phi(\boldsymbol{x})$의 미분이 요구된다. 식(6)의 에너지 함수는 아래와 같은 미분값을 갖는다.

$$\nabla_{\boldsymbol{x}}G_\phi(\boldsymbol{x})=\boldsymbol{x}-\nabla_{\boldsymbol{x}}^T\boldsymbol{f}_\theta(\boldsymbol{x})\boldsymbol{f}_\theta(\boldsymbol{x}) \tag{7}$$

여기서 $\nabla_{\boldsymbol{x}}\boldsymbol{f}_\theta(\boldsymbol{x})$은 Jacobian 행렬, T는 전치(transpose)를 나타낸다. 식(7)을 기반으로 본 논문에서는 계산량 감소를 위해 CD-1 알고리즘[4]과 비슷하게 학습 데이터 $\boldsymbol{x}_0$를 초기값으로 하는 간단한 1-step SGLD을 사용한다.

$$\boldsymbol{x}=\nabla_{\boldsymbol{x}}^T\boldsymbol{f}_\theta(\boldsymbol{x}_0)\boldsymbol{f}_\theta(\boldsymbol{x}_0)+\gamma\epsilon,\ \epsilon\sim N(0,I) \tag{8}$$

위 식은 식(2)에 식(7)를 대입하고 $\alpha_t=1$, $\gamma=\sqrt{2\alpha_t}$로 설정하여 얻게 된다. 실험을 통해 위와 같은 샘플링 과정이 매우 잘 동작하는 것을 알 수 있다.

제안된 CEBM에 대한 NCE 학습 알고리즘은 다음과 같다.

제안된 CEBM의 NCE 학습 알고리즘

입력: 에너지 함수 $D_\theta(\boldsymbol{x}), G_\phi(\boldsymbol{x})$, epoch T, batch size n, 학습률 η, 매개변수 γ

출력: 최적의 파라미터 값, θ^*,ϕ^*

1: θ,ϕ 값 초기화

2: for $t=0:T$ do

3: sample $\boldsymbol{x}_1,\boldsymbol{x}_{2,\ldots,}\boldsymbol{x}_n$ from data

4: sample $\epsilon_1,\epsilon_2,\ldots,\epsilon_n$ from $\epsilon\sim N(0,I)$

$\boldsymbol{y}_i=\nabla_{\boldsymbol{x}}^T\boldsymbol{f}_\phi(\boldsymbol{x}_i)\boldsymbol{f}_\phi(\boldsymbol{x}_i)+\gamma\epsilon_i$ 계산

5 :

$V(\theta)=-\frac{1}{n}\sum_i\left(\log\sigma(-D_\theta(\boldsymbol{x}_i))+\log(1-\sigma(-D_\theta(\boldsymbol{y}_i)))\right)$

6: $\theta\leftarrow\theta-\eta\nabla_\theta V(\theta)$

7: $V(\phi)=\frac{1}{n}\sum_i\log(1-\sigma(-D_\theta(\boldsymbol{y}_i)))$

8: $\phi\leftarrow\phi-\eta\nabla_\phi V(\phi)$

9: end for

Ⅲ. 실 험

이 장에서는 제안된 CEBM의 성능 평가를 위해 준지도 기반 이상탐지 실험을 수행하였다. 이상탐지를 위해 3가지 데이터 세트, ECG, UNSW 그리고 MNIST/Fashion-MNIST를 이용하였다. AE, ML 기반 EBM, SM 기반 EBM 그리고 기타 모델 등을 사용하여 성능을 비교하였다. 이상탐지 성능 평가의 척도로 F1-score[8]를 이용하였다.

3.1 ECG

ECG5000[19]은 20시간 길이의 심전도 데이터로 각 하트 비트를 추출하고 보간법을 사용하여 각각의 길이를 같게 전처리하였다. 심전도 전체 데이터 수는 5,000개이며, 각 데이터는 140개의 실수 값으로 구성된다. 모든 데이터는 [0,1] 사이로 minmax 정규화되었다. 각 심전도는 1(비정상) 또는 0(정상)으로 레이블링되고 학습에는 정상 1639개, 테스트에는 정상/비정상 1000개를 사용하였다.

CEBM의 훈련을 위해 학습률 $\eta=0.0001$을 갖는 Adam optimizer을 이용하였고, batch size 100, 그리고 SGLD 파라미터로 $\gamma=0.1$를 설정하였다. Epoch은 최고의 성능이 나오도록 선택되었다. CEBM의 두 에너지 함수 $D_\theta(\boldsymbol{x})$와 $G_\phi(\boldsymbol{x})$을 구성하는 신경망은 완전연결신경망(Fully-Connected NN, FCNN)을 사용하였다. $D_\theta(\boldsymbol{x})$와 $G_\phi(\boldsymbol{x})$의 신경망은 같은 층으로 구성하였고,

3층의 경우에 각각의 노드 수는 (140-1024-1024-1)와 (140-1024-1024-512)를 사용하였다.

먼저 CEBM의 이상탐지 특성을 파악하기 위해 정상과 비정상 데이터에 대한 $D_\theta(\boldsymbol{x})$와 $G_\phi(\boldsymbol{x})$ 값에 대한 히스토그램을 사용하였다. 그림 1은 $D_\theta(\boldsymbol{x})$(위)와 $G_\phi(\boldsymbol{x})$(아래) 값에 대한 정상과 비정상 데이터의 히스토그램과 최적의 임계값을 나타낸다. 이때 $D_\theta(\boldsymbol{x})$의 평균과 분산은 정상은 (5.2, 1.1), 비정상은 (2.7, 0.5)이며, 최적의 임계값은 3.86이다. $G_\phi(\boldsymbol{x})$는 각각 (-8.876, 0.027), (-8.907, 0.016) 그리고 -8.88이다. 그림 1에서 정상과 비정상의 $D_\theta(\boldsymbol{x})$ 값들은 변별력 있게 분포하지만, $G_\phi(\boldsymbol{x})$는 많은 부분에서 중첩해서 나타났다. 이는 $D_\theta(\boldsymbol{x})$가 데이터를 잡음분포와 대비하여 변별력 있게 학습할 결과라고 할 수 있다. 따라서 CEBM의 성능평가 척도로 $D_\theta(\boldsymbol{x})$ 값을 임계값과 비교하여 정상과 비정상을 판별하였다. 임계값은 훈련 데이터의 $D_\theta(\boldsymbol{x})$의 평균값을 중심으로 최적의 값을 선택하였다.

그림 2는 ML와 SM 기반 EBM 그리고 제안된 NCE 기반 CEBM에 대한 은닉층 수에 따른 성능을 나타낸다. 3층일 때 모든 기법이 가장 높은 성능을 나타내었고, CEBM은 2층과 3층에서 비슷한 결과를 보였다. 제안된 NCE 기반 CEBM은 ML과 SM보다 모든 층에서 더 높은 성능을 나타내었다. 표 1에서 제안된 기법과 AE, VAE 그리고 transformer 기법의 성능을 추가로 비교하였다. 표에서 살펴보듯이 제안된 기법은 AE와 VAE[23]보다 높은 96.53%의 F1-score를 나타내었다. 이때 CEBM의 precision과 recall은 각각 93.62%와 99.64%를 기록하였다. 그러나 최근에 제안된 더 복잡한 구조를

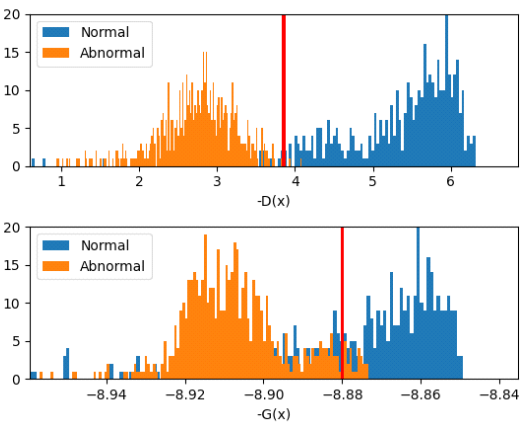

그림 1. CEBM의 두 에너지 함수에 대한 히스토그램, (위) $D_\theta(\boldsymbol{x})$ (아래) $G_\phi(\boldsymbol{x})$
Fig. 1. Histogram of the two energy functions of CEBM, (Top) $D_\theta(\boldsymbol{x})$ (Bottom) $G_\phi(\boldsymbol{x})$.

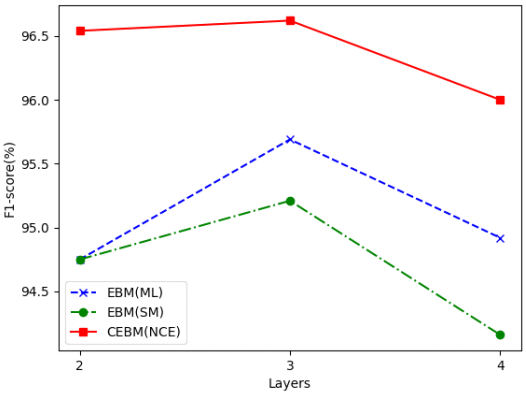

그림 2. ECG에서 은닉층 수에 따른 EBM의 학습기법에 대한 F1-score (%)
Fig. 2. The F1-score(%) on the learning methods of EBM according to the number of hidden layers in ECG.

표 1. ECG에서 NCE 기반 CEBM와 다른 기법들과의 이상탐지 성능비교
Table 1. Comparison of anomaly detection performance with NCE-based CEBM and other techniques in ECG.

Models	F1-score (%)
AE[7]	92.90
EBM(ML)[7]	95.69
EBM(SM)[7]	95.21
VAE[23]	96.01
Transformer[24]	99.00
CEBM(NCE)	**96.53** (Precision/Recall: 93.62/99.64)

갖는 transformer[24]에 비해 더 낮은 성능을 나타내었다. 따라서 ECG 실험에서 제안된 기법이 최근 모델인 transformer 기법에 비해 더 낮지만, 비슷한 구조를 갖는 EBM과 AE/VAE 기법들에 비해 이상탐지에 더 효과적임을 알 수 있다.

3.2 UNSW

UNSW NB15[20]는 최신 네트워크 트래픽 특성 과 공격 기법을 반영한 네트워크 침입탐지를 위한 대표적인 데이터 셋이다. 이 데이터는 여러 개의 트래픽 생성 알고리즘을 통해 만들어진 가상의 네트워크 트래픽을 캡처하여 생성되었다. 데이터는 네트워크에서 발생할 수 있는 9개의 공격 유형과 1개의 정상 유형을 가지는 레이블로 구성되었다. 이상탐지를 위해 9개의 공격 유형은 '1'로, 정상은 '0'으로 표시되었다. 각 네트워크 데이터는 42개의 특성을 포함한다. 42개의 특성은 네트워크의 특징을 나타내는 연속적인 값과 이산적인 값으로 구성되었다. 이 특성들을 신경망의 입력으로 사용하

표 2. UNSW에서 NCE 기반 CEBM와 다른 기법들과의 이상탐지 성능비교
Table 2. Comparison of anomaly detection performance with NCE-based CEBM and other techniques in UNSW.

Models	F1-score (%)
AE[7]	78.92
EBM(ML)[7]	80.10
EBM(SM)[7]	80.36
FLAE[25]	83.63
CEBM(NCE)	**87.14** (Precision/Recall: 95.21/80.34)

기 위해 188개의 특징 벡터값을 갖도록 변환되었다. 최종적으로 모든 데이터에 대해 [0,1] 사이로 minmax 정규화로 전처리되었다. 학습에는 정상 56,000개, 테스트는 82,332개(정상: 37,000, 비정상: 45,332)를 사용하였다. 학습을 위한 하이퍼파라미터는 ECG와 비슷하게 설정되었다. $D_\theta(\boldsymbol{x})$와 $G_\phi(\boldsymbol{x})$의 신경망은 3층으로 구성하였고, (188-512-512-1)와 (188-512-512-256)의 노드 수를 각각 사용하였다. 성능 비교를 위해 AE, ML 및 SM EBM 그리고 최근에 제안된 침입탐지를 위한 federated learning with autoencoder (FLAE)[25]와 비교하였다.

표 2는 UNSW에서 제안된 기법과 다른 기법들과의 이상탐지 성능을 비교하였다. ECG와 마찬가지로 AE, ML 및 SM 기반 EBM 보다 제안된 모델이 더 높은 F1-score를 나타내었다. 또한 최근에 제안된 FLAE의 83.63%보다 더 높은 86.73%의 성능을 나타내었다. 이때 CEBM의 precision과 recall은 각각 96.10%와 79.02%를 나타내었다. 따라서 UNSW 데이터 셋과 같은 네트워크 이상탐지 실험에서도 제안된 기법이 매우 효과적임을 알 수 있다.

3.3 MNIST/Fashion-MNIST

MNIST[21]와 Fashion-MNIST[22]는 28x28 크기의 gray 이미지로 각각 10개의 클래스를 갖는 데이터 세트이다. MNIST를 정상 데이터로, Fashion MNIST를 이상 데이터로 하는 이미지에 대한 이상탐지 실험을 수행하였다. 학습에는 정상으로 MNIST 60,000개의 훈련 데이터를, 테스트에는 정상과 비정상으로 각각 MNIST와 Fashion-MNIST 테스트 데이터 셋 10,000개씩을 사용하였다.

학습을 위한 하이퍼파라미터는 ECG와 같고, SGLD 파라미터로 $\gamma = 0.01$를 사용하였다. CEBM의 $D_\theta(\boldsymbol{x})$와 $G_\phi(\boldsymbol{x})$의 신경망은 3층 FCNN 형태로, 각각 (784-1024-1024-1)와 (784-1024-1024-512)를 사용하였다. 표 3은 제안된 CEBM과 다른 기법들과의 성능을 나타낸다. AE에 비해 ML과 SM 기반 EBM이 더 높은 성능을 나타내었고, ML EBM이 98.38%로 SM보다 더 뛰어난 성능을 보였다. 제안된 CEBM은 99.32%로 다른 기법들에 비해 더 뛰어난 F1-score를 나타내었다. 이때 precision/recall은 98.91%/99.74%이다. 따라서 MNIST/Fashion-MNIST와 같은 이미지 이상탐지 실험에서도 제안된 CEBM이 다른 기법에 비해 더 효과적임을 알 수 있다.

표 3. MNIST/Fashion-MNIST에서 NCE 기반 CEBM와 다른 기법들과의 이상탐지 성능비교
Table 3. Comparison of anomaly detection performance with NCE-based CEBM and other techniques in MNIST/Fashion-MNIST.

Models	F1-score (%)
AE[7]	86.17
EBM(ML)[7]	98.38
EBM(SM)[7]	95.36
CEBM(NCE)	**99.32** (Precision/Recall: 98.91/99.74)

Ⅳ. 결 론

본 논문에서는 두 개의 비정규화된 에너지 함수로 구성된 확률분포의 곱으로 새로운 CEBM 구조를 제안하였다. CEBM을 NCE 기법으로 학습하기 위해 하나의 분포를 잡음분포로 사용하였고 이를 위한 목적함수를 유도하였다. 제안된 목적함수를 위해 잡음분포로부터 데이터를 생성하기 위한 근사적인 1-step SGLD에 기반한 샘플링 기법을 제시하였다. 제안된 NCE 기반 CEBM의 성능을 평가하기 위해 3가지 데이터 세트, ECG, UNSW 그리고 MNIST/Fashion-MNIST에 대한 이상탐지 실험을 수행하였다. 모든 데이터 세트에서 기존의 AE, ML과 SM 기반의 EBM에 비해 제안된 CEBM이 더 높은 F1-score를 나타내었다. 따라서 제안된 CEBM 기반 NCE 학습방법은 이상탐지에 매우 효과적임을 나타내었다.

향후 연구 방향은 먼저 대용량의 고차원 데이터에 대한 CEBM의 특성을 파악하고 이를 이상탐지에 적용하는 것이 필요하다. 그리고 텍스트와 같이 순차적인 특징을 갖는 데이터에 대한 이상탐지를 위해 FCNN 대신 transformer와 같은 구조를 포함한 CEBM에 대한 연구가 필요하다.

References

[1] Y. LeCun, S. Chopra, R. Hadsell, M. Ranzato, and F. Huang, "A tutorial on energy-based learning," *Predicting Structured Data,* Jan. 2006.

[2] Y. Song and D. P. Kingma, "How to train your energy-based models," *arXiv preprint arXiv:2101.03288,* 2021. (https://doi.org/10.48550/arXiv.2101.03288)

[3] J. Ngiam, Z. Chen, P. W. Koh, and A. Y. Ng, "Learning deep energy models," in *Proc. 28th ICML*, pp. 1105-1112, 2011.

[4] G. E. Hinton, "A practical guide to training restricted Boltzmann machines," *Tech. Rep. UTML TR2010-003*, University of Toronto, 2010. (https://doi.org/10.1007/978-3-642-35289-8_32)

[5] M. M. Al Rahhal, Y. Bazi, R. Al-Dayil, B. M. Alwadei, N. Ammour, and N. Alajlan, "Energy-based learning for open-set classification in remote sensing imagery," *IJRS*, vol. 43, no. 15-16, pp. 6027-6037, 2022. (https://doi.org/10.1080/01431161.2022.2044539)

[6] P. Guo and D. Kim, "A new energy-based latent-variable model for unsupervised feature learning," *J. KICS*, vol. 48, no. 5, pp. 509-516, 2023. (https://doi.org/10.7840/kics.2023.48.5.509)

[7] P. Guo and D. Kim, "A multi-layer hopfield neural network for semi-supervised anomaly detection," *J. Digital Contents Soc.*, vol. 24, no. 11, pp. 2893-2900, Nov. 2023.

[8] R. Chalapathy and S. Chawla, "Deep learning for anomaly detection: A survey," *arXiv preprint arXiv:1901.03407*, Jan. 2019. (https://doi.org/10.48550/arXiv.1901.03407)

[9] H. Estiri and S. N. Murphy, "Semi-supervised encoding for outlier detection in clinical observation data," *Computer Methods and Programs in Biomedicine*, vol. 181, Nov. 2019. (https://doi.org/10.1016/j.cmpb.2019.01.002)

[10] M. Kliger and S. Fleishman, "Novelty detection with GAN," *arXiv preprint arXiv:1802.10560*, Feb. 2018. (https://doi.org/10.48550/arXiv.1802.10560)

[11] J. Besag, "Comments on representations of knowledge in complex systems," by U. Grenander and Mi Miller, *J. Roy. Statist. Soc. Ser.* B, vol. 56, pp. 591-592, 1994.

[12] A. Hyvärinen, "Estimation of non-normalized statistical models by score matching," *J. Machine Learn. Res.*, vol. 6, pp. 695-709, 2005.

[13] K. Swersky, M. A. Ranzato, D. Buchman, B. M. Marlin, and N. D. Freitas, "On autoencoders and score matching for energy based models," in *Proc. 28th ICML-11*, pp. 1201-1208, 2011.

[14] D. P. Kingma, "Improving score matching for learning statistical models of natural images," Ph.D. dissertation, New York University, 2020.

[15] M. Gutmann and A. Hyvärinen, "Noise-contrastive estimation: A new estimation principle for unnormalized statistical models," *PMLR*, pp. 297-304, May 2010.

[16] C. Ceylan and M. Gutmann, "Conditional noise-contrastive estimation of unnormalised models," in *Int. Conf. Machine Learn.*, pp. 726-734, 2018.

[17] A. Bose, H. Ling, and Y. Cao, "Adversarial contrastive estimation," in *Proc. 56th Annual Meeting of the Assoc. for Computational Linguistics,* vol. 1, pp. 1021-1032, 2018.

[18] R. Gao, et al., "Flow contrastive estimation of energy-based models," in *Proc. IEEE/CVF Conf. CVPR*, Jun. 2020.

[19] S. K. Berkaya, A. K. Uysal, E. S. Gunal, S. Ergin, S. Gunal, and M. B. Gulmezoglu, "A survey on ECG analysis," *Biomedical Signal Process. and Control,* vol. 43, pp. 216-235, May 2018. (https://doi.org/10.1016/j.bspc.2018.03.003)

[20] N. Moustafa and J. Slay, "UNSW-NB15: A comprehensive data set for network intrusion detection systems (UNSW-NB15 network data set)," in *Proc. MilCIS*, pp. 1-6, Canberra, Australia, Nov. 2015. (https://doi.org/10.1109/MilCIS.2015.7348942)

[21] AT & T Labs, MNIST handwritten digit database [Internet] Available: http://yann.lecun.com/exdb/mnist

[22] H. Xiao, K. Rasul, and 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)

[23] P. Matias, D. Folgado, H. Gamboa, and A. V. Carreiro, "Robust anomaly detection in time series through variational autoencoders and a local similarity score," in *Proc. 14th Int. Joint Conf. Biomedical Eng. Syst. and Technol. (BIOSTEC)*, Online, pp. 91-102, Feb. 2021. (https://doi.org/10.5220/0010320500002865)

[24] A. Alamr and A. Artoli, "Unsupervised transformer-based anomaly detection in ECG signals," *Algorithms*, vol. 16, no. 3, Mar. 2023. (https://doi.org/10.3390/a16030152)

[25] O. Aouedi, K. Piamrat, G. Muller, and K. Singh, "Intrusion detection for softwarized networks with semi-supervised federated learning," in *Proc. IEEE ICC*, pp. 5244-5249, Seoul, May 2022. (https://doi.org/10.1109/ICC45855.2022.9839042)

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 : 삼성전자 전문연구원

2003년 4월~2004년 2월 : 한국전자통신연구원 선임연구원

2004년 2월~현재 : 전남대학교 전자공학과 교수

<관심분야> 딥러닝, 기계학습, 인공지능신호처리

[ORCID:0000-0001-9316-7069]