

# 지식 증류를 통한 자가 개선 BeatGAN

임에딘1 황채은2 박경문3

1,2,3경희대학교 컴퓨터공학과

vlfdu0401@khu.ac.kr, hce9603@khu.ac.kr, gmpark@khu.ac.kr

## Self-Improving BeatGAN via Knowledge Distillation

Edin Lim1 Chae-eun Hwang2 Gyeong-Moon Park3

1,2,3Computer Science and Engineering, Kyung Hee University

### 요 약

최근 사람들의 건강에 대한 관심사가 늘고 심장 질환에 대한 걱정이 커짐에 따라, 병원에서 부정맥을 진단하고 치료하는 것에 대한 중요도도 커지고 있다. 현재 병원에서는 환자들의 심전도 데이터를 측정 및 기록하고 분석하는 일을 전문의에게 의존하고 있으며, 이는 많은 수고로움을 동반한다. 따라서 본 연구에서는 그러한 문제점을 개선하고자 기존의 비지도 학습 기반 모델을 사용하여 비정상 심전도를 감지하는 BeatGAN에 지식 증류 기법을 도입하여 자가 개선을 하는 부정맥 예측 딥러닝 모델을 제안한다.

## 1. 서 론

최근 기술과 의학이 발전함에 따라, 사람의 평균 수명은 예전보다 많이 늘어났다. 이에 사람들은 늘어난 수명만큼 오랜 시간을 건강하게 보내려고 노력하고, 당연히 자신과 가족의 건강 문제에 대한 관심 또한 늘어나고 있다.

이러한 관심이 늘어나면서, 최근 들어 많은 사람에게 자주 발생하고 막대한 영향을 주는 질병 중 하나인 심장 및 혈관 질환의 관찰 및 예측은 중요도가 높아지게 되었다. 심장 질환 및 혈관 질환은 아직도 계속 늘어나는 추세이며, 그 중에서도 자주 언급되는 질환 중 하나가 바로 부정맥이다. 부정맥이란, 심장이 불규칙한 리듬이나 비정상적인 심박수를 갖는 것으로, 돌연사나 뇌졸중 등 생과 직결된 중질환을 유발할 수 있다. 부정맥은 치료가 아예 불가능한 질병은 아니나, 치료하기 위해서는 정확한 진단이 필요하다. 부정맥 진단의 가장 보편적인 방법은 심전도 검사이며, 이를 통해 나온 심전도 데이터(ECG 데이터)를 통해 의사는 부정맥을 진단하고 치료를 시행할 수 있다.

우리는 ECG 데이터에서 비정상적인 리듬을 감지하여 부정맥을 판단할 수 있는 시 기반의 딥러닝 알고리즘인 BeatGAN에 지식 증류 기법을 융합하여 자가 개선을 하는 딥러닝 모델을 제안한다[1,2].

본 논문의 구성은 다음과 같다. 2장에서는 BeatGAN과 지식 증류에 대해 설명을 하고, 3장에서는 제안된 모델의 네트워크 구조와 최적화 목표를 제시한다. 4장에서는 실험을 통해 기존의 BeatGAN과 지식 증류를 활용하여 자가 개선을 하는 BeatGAN의 성능을 비교한다. 마지막으로 5장에서는 결론을 맺는다.

## 2. 관련 연구

### 2.1. BeatGAN

생성자(Generator)와 구분자(Discriminator) 두 네트워크를 활용하여 적대적으로 생성된 시계열 데이터를 통해 ECG 데이터에서 비정상적인 리듬을 감지하는 알고리즘이다[1]. (BeatGAN: Anomalous Rhythm Detection using Adversarially Generated Time Series) BeatGAN은 다음과 같은 장점을 가진다. 1) 비지도 방식(Unsupervised) : 레이블 없이 적용이 가능하다. 2) 효율성(Effectiveness) : ECG 데이터에서 거의 0.95 AUC의 정확도를 달성하고 매우 빠른 추론 속도(비트당 2.6ms)를 보인다. 3) 설명 가능성(Explainability): 비정상적인 패턴과 그와 관련된 시간 턱을 정확히 찾아내어 시각화 및 주의 집중을 위한 해석 가능한 출력을 제공한다. 4) 일반성(Generality) : ECG 데이터와 CMU Motion Capture 데이터에서 비정상적인 움직임을 성공적으로 감지한다.

#### 2.1.1. 이상 탐지 기술(Anomaly Detection)

정상 시스템에서 발생하는 의도치 않은 동작을 이상 현상이라고 하며, 이를 해결하기 위해서는 발생한 상황이 ‘이상 상태’임을 인지해야 한다. 이상 현상은 정상 상태와 형이하여 쉽게 인지할 수도 있지만, 정상 상태와 매우 유사하여 인지하기 어려울 때도 있다. 이를 탐지하기 위해 ‘이상 탐지 기술’ 연구가 진행되고 있다. 이상 탐지 기술은 학습 시 비정상 샘플과 레이블의 유무에 따라 지도(Supervised) 이상 탐지와 비지도(Unsupervised) 이상 탐지로 나뉜다. BeatGAN의

경우 비지도 이상 탐지 기술을 활용하기 때문에 라벨링 과정이 필요 없다는 장점이 있으나, 다른 방법에 비해 정확도가 낮다는 단점을 가진다.

### 2.1.2. 재구성 판별 방식(Reconstruction)

재구성 기반 판별 방식(Reconstruction)은 이상을 판별할 데이터를 저차원 형태의 구조(Latent structure)를 획득하고, 그 이후 인위적으로 재구성한 데이터를 생성하기 위한 모델을 사용한다. 잠재 구조를 활용하는 이유는 이상(anomaly) 현상을 저차원으로 매핑하면 정보가 손실되어 효과적으로 재구성할 수 있기 때문이다. 고차원 데이터를 저차원 데이터로 전환하는 대표적인 방법으로는 PCA(Principal Component Analysis, 주성분 분석)이 있으나, 선형 재구성으로 제한되며 상관 관계가 높은 분포에만 적용이 가능하다는 단점이 있다. 딥러닝 기반 기술에서는 주로 오토인코더(Autoencoder) 기반의 방법론이 자주 사용된다. PCA와 달리 비선형 차원 축소를 다룰 수 있다.

### 2.2. 지식 증류(Knowledge Distillation)

지식 증류(Knowledge Distillation)는 큰 모델(Teacher Network)로부터 작은 모델(Student Network)로 지식을 증류시켜 전달하는 방식이다. 복잡하고 큰 모델은 실제 사용하기에 부적합할 수 있고 자원을 많이 사용하므로 학습을 통해 필요한 정보만을 남긴 작은 모델을 활용함으로써 성능을 높이는 것을 목적으로 한다. 지식 증류는 softmax 함수에 매개변수  $T$ (Temperature)을 반영함으로써 출력값을 soft label로 하여 모방 학습에서의 이점을 최대화하고 성능을 높인다.

## 3. 모델 구성 및 설계

### 3.1. BeatGAN 네트워크 구조

BeatGAN 모델의 네트워크 구조는 그림 1에 묘사되어 있다. 오토인코더는 인코더  $G_E$ 와 디코더  $G_D$ 로 구성되어 있으며, 입력 데이터  $x$ 가 인코더로 들어오면 은닉 벡터  $z$ 로 압축되고  $G_D(z)$ 는  $x'$ 를 생성한다. 여기에 적대적인 학습 방식인 GAN 기술을 추가하여 Discriminator( $D$ )가 정규화의 역할을 수행한다.

디코더 네트워크인  $G_D$ 의 구조는 DCGAN의 생성자 구조와 비슷하다[3]. DCGAN은 기존 GAN에 존재했던 완전 연결된 구조의 대부분을 CNN구조로 대체한 것이다.  $G_D$ 는 그림 2와 같이 1D transposed 컨볼루션 레이어들과 Batch-norm, Leaky ReLU로 구성되어 있다.  $G_E$ 는  $G_D$ 와 거의 유사하며 반대 방향으로 되어있다.

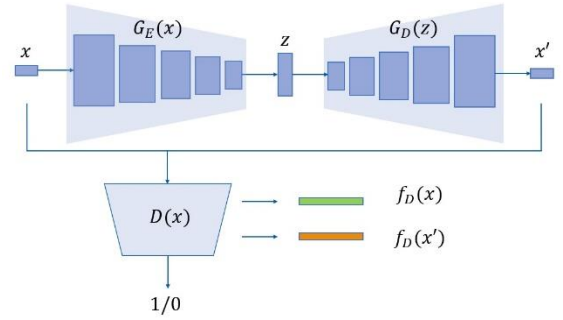


그림 1. BeatGAN 네트워크 구조

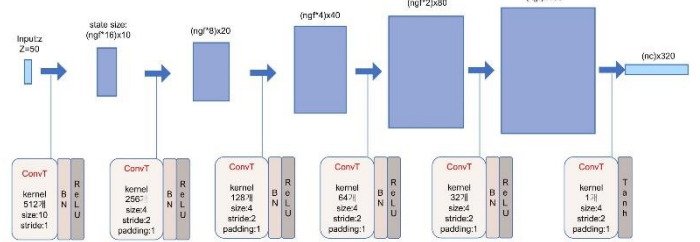


그림 2. 디코더 네트워크  $G_D$  구조

### 3.2. 지식 증류를 통한 자가 개선 BeatGAN 모델

우리는 지식 증류 기법을 활용하기 위해 교사 네트워크와 학생 네트워크를 설정했다. 제안하는 모델의 구조는 그림 3에 묘사되어 있다. 기존 BeatGAN 모델을 사전에 훈련시키고 이를 교사 네트워크로 삼아서, 데이터에 대해 예측한 결과를 학생 네트워크가 전달받는다. 학생 네트워크는 훈련하는 과정에서 (1) 적대적 정규화를 통한 재구성의 손실, (2) 학생이 예측한 결과와 교사로부터 전달받은 결과와의 손실을 통합하여 업데이트를 진행한다.

기존의 BeatGAN에서 재구성 모델의 손실 함수는 다음과 같으며 이를 최소화하는 것이 목적이다.

$$L_G = ||x - x'||_2 + \lambda ||f_D(x) - f_D(x')||_2 \quad (1)$$

왼쪽 항은  $G$ 를 통해 생성된  $x'$ 와  $x$ 의 복원 오차(reconstruction error)이다. 이는 이상 점수(anomaly score)가 되어 threshold와 비교를 통해 이상 여부를 결정한다. 오른쪽 항은 특징 일치 손실(feature matching loss)로, 손실을 계산하기 위해  $D$ 의 은닉층의 활성화 벡터  $f_D(\cdot)$ 을 이용하여 생성된  $x'$ 와  $x$ 의 특징 사이가 유사한지를 고려한다.

다른 지식 증류 기법과 유사하게 본 연구에서도 소프트 레이블을 통해 증류를 수행한다. 학생 네트워크가 출력하는 예측 결과와 교사 네트워크가 예측하는 결과에 대해 쿨백-라이블러 발산(KL divergence)을 통해 다음과 같이 손실 함수를 정의한다.

$$L_{KD}(x; \theta_s, \theta_t, T) \quad (2)$$

$$= D_{KL}(\text{softmax}(\frac{f_s(x; \theta_s)}{T}) || \text{softmax}(\frac{f_t(x; \theta_t)}{T}))$$

$T$ 는 온도 매개변수,  $s$ 와  $t$ 는 각각 학생과 교사 모델

매개변수이다. 우리는 위의 손실함수를 통합하여 다음과 같은 최적화 목적을 구성했다.

$$L = L_G + L_{KD}(x; \theta_s, \theta_t, T) \tag{3}$$

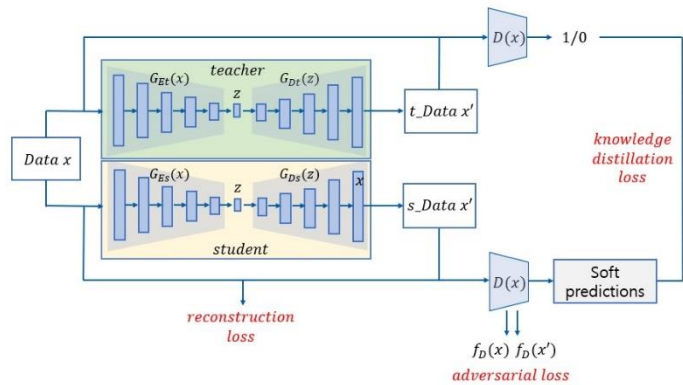


그림 3. 자가 개선 BeatGAN 구조

## 4. 연구 및 결과

### 4.1. 데이터

데이터는 MIT-BIH arrhythmia database를 이용했다[4]. 일반적으로 부정맥 감지를 평가하는 데 사용할 수 있는 데이터로, 심전도와 관련있는 딥러닝 분야에서 많이 사용되고 있다. 이 데이터에는 피험자 47명으로부터 수집한 2채널 구성의 24시간 ECG 기록을 30분동안 발췌하여 총 48개로 구성되어 있으며, 본 논문의 연구에서 기존 BeatGAN의 연구에 이용하던 데이터를 사용하였다. 총 데이터에는 97,568개의 비트와 2,860만의 시간 틱이 포함되어 있다.

### 4.2. 연구 결과

표 1은 기존의 BeatGAN 모델로 사전에 훈련을 시킨 후에 테스트한 결과와, 사전 훈련된 교사 네트워크를 이용하여 지식증류 기법으로 학생 네트워크를 훈련시킨 후 테스트한 결과를 표로 나타낸 것이다. 표 1에서 볼 수 있듯이 기존의 모델보다 자가 개선을 수행하는 SI-BeatGAN의 정확도가 향상되었음을 알 수 있다. 그림 4는 BeatGAN과 SI-BeatGAN의 ROC 비교 그래프이며 마찬가지로 우리가 제안한 모델의 정확도가 더 높은 결과를 보여주고 있다.

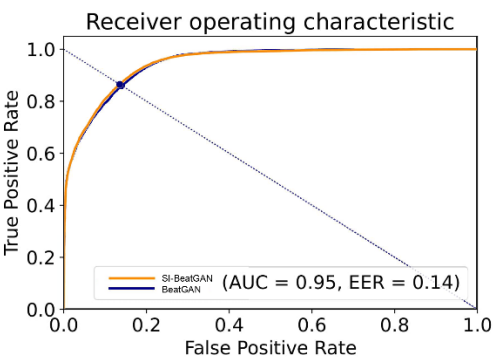


그림 4. ROC 비교 그래프

## 5. 결론

본 논문에서는 기존의 BeatGAN에서 교사 네트워크와 학생 네트워크 구성을 하고 지식 증류 기법을 융합하여 자가 개선을 하는 SI-BeatGAN을 소개하였다. 학생 네트워크는 훈련하는 과정에서 학습된 교사 네트워크의 지식을 전달받아 자가 개선을 수행하였으며 그 결과 기존 모델보다 더 높은 정확도를 보여주었다.

### 참 고 문 헌

[1]Zhou B, Liu S, Hooi B, Cheng X, Ye J. BeatGAN: Anomalous Rhythm Detection using Adversarially Generated Time Series, Proceedings of the Twenty-Eighth International Joint Conference on Artificial Intelligence Main track. Pages 4433–4439, 2019  
[2]Ji M, Shin S, Hwang S, Park G, Moon I. Refine Myself by Teaching Myself : Feature Refinement via Self-Knowledge Distillation. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), pp. 10664–10673, 2021  
[3] Alec Radford, Luke Metz, and Soumith Chintala. Unsupervised representation learning with deep convolutional generative adversarial networks. arXiv preprint arXiv:1511.06434, 2015  
[4] George B Moody and Roger G Mark. The impact of the mit-bih arrhythmia database. IEEE Engineering in Medicine and Biology Magazine, 20(3):45–50, 2001

표 1. ECG 데이터에서 BeatGAN과 지식증류가 수행된 SI-BeatGAN의 성능 차이

방법	AUC	AP
BeatGAN	약 0.945893	약 0.910805
SI-BeatGAN	약 0.947469	약 0.914809