

# 설명 가능한 딥 러닝을 이용한 심전도 데이터 분석 연구

유재상<sup>o</sup>, 김성태

경희대학교 컴퓨터공학과

jsyoo1996@khu.ac.kr st.kim@khu.ac.kr

## Electrocardiogram data analysis study using explainable deep learning

Jaesang Yoo<sup>o</sup>, Seongtae Kim

Department of Computer Science and Engineering, Kyung Hee University

### 요약

최근 AI 기술이 4 차 산업혁명의 핵심기술로 자리잡으면서, 관련 연구가 활발하게 진행되고 있다. 특히 기계학습분야에 많은 발전이 있었고, 다양한 모델들이 다양한 분야에서 실험적으로, 또 실용적으로 사용되고 있다. 의료분야에서도 쓰이고 있는데 심전도 데이터(ECG) 분석의 경우가 대표적이다. 기존에는 숙달된 인력이 여러 파형을 보고 분석해 결과를 추론해야 했지만, 딥 러닝 모델을 통해 이 과정을 단순화, 가속화할 수 있음과 동시에 높은 정확성을 갖게 한다. 하지만 동시에 한계점을 가지는데, 바로 기계학습 모델의 블랙박스 특성으로 인해 어떤 과정을 통해 결론을 도출하였는지 알기 어렵다는 것이다. 때문에 이에 관한 연구를 통해 결과의 신뢰성을 높이고, 판별을 위해 기존에 사용되지 않았던 중요한 특징을 역으로 발견할 수도 있다.

### 1. 서론

딥 러닝 모델의 높은 정확도와 가능성으로 인해 쓰이지 않는 곳을 찾아보기가 더 어려워졌다. 이는 기존에 인간이 많은 시간과 정성을 들여서 해야 했던 일을 쉽게 대신 해내기 시작했고, 그를 넘어서 인간이 해내지 못했던 분야까지 해내고 있다. 하지만 모델이 점점 더 복잡해짐에 따라 블랙박스라는 특성을 갖게 되는데, 결과의 정확도는 높지만, 어떻게 그러한 결과를 도출했는지는 들여다보기 어렵게 된 것이다.

이 맹점은 실패했을 경우의 리스크가 큰 분야일수록 크게 다가온다. 특히나 의료분야에서 활발히 연구되고 있는 ECG 분석의 경우가 그렇다. 심전도 데이터로 크게는 부정맥이나 빈혈의 가능성을 진단할 수 있는데, 실패할

경우 인명에 영향을 미친다. 또한, 환자의 경우에는 증상에 대해 설명 받을 권리가 있고, 의사의 경우 설명해주어야 할 의무가 존재하는데, 블랙박스 형태의 모델의 경우 여기서 벗어나 있다.

따라서 단순히 모델의 높은 정확도만을 목표로 할 것이 아니라, 동시에 특이점을 포착해서 실질적인 사용자(모델의 개발자가 아닌)가 이해하기 쉬운 정보로 표현할 수 있어야 한다. 이 과정이 잘 정착되면, 사용의 용이성뿐만 아니라, 딥 러닝 모델의 신뢰성을 확보할 수 있는 장치로 자리매김할 것이다.

정상적인 ECG 데이터와 비정상적이거나 증상의 분류가 되어있는 ECG 데이터를 다양한 딥러닝 모델에 학습시킨 후, 환자의 ECG 데이터를 입력하면 건강의

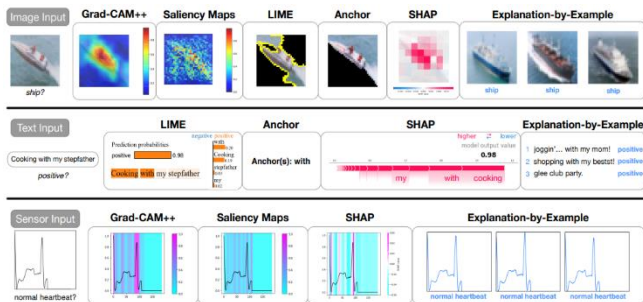
이상유무를 파악하고 질환자의 ECG 데이터도 충분히 학습되었다면 진단까지 가능하다. 또한, 모델의 판단 근거를 정상 파형과 환자의 파형을 출력하여 의사와 환자가 이해하고 납득할 수 있도록 하고, 더 나아가서는 해당 출력을 이용해 의사가 특이사항 판별법을 역으로 발견하는 계기가 될 수 있다.

## 2. 관련연구

### 2.1. 딥러닝 모델의 설명 방법

서론에서 언급했듯 딥러닝 모델이 시간이 지날수록 높은 정확도를 가지게 되었으나, 동시에 모델이 어떤 특징을 포착해 판단을 냈는지 그 신뢰성은 떨어지게 되었다. 모델의 구조가 복잡해 사용자가 직관적으로 알 수 없게 된 것이다. 때문에 자연스럽게 딥러닝 모델의 설명 방법에 관한 연구도 이루어지게 되었다.

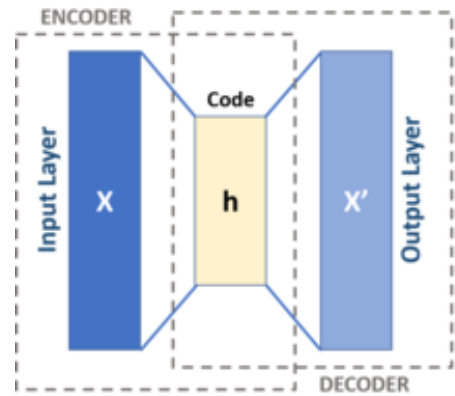
간단한 방법으로는 훈련 데이터 중 가장 인접한 데이터 값을 표현하는 방법부터, 훈련 데이터의 분포 중첩도를 그래프 형태로 출력해주는 방법 등, 다양한 데이터들에 따라 적합한 설명 모델들이 연구되고 있다.



[그림 1] 다양한 데이터 타입과 그에 따른 설명 방법

### 2.2.1 Autoencoder

Autoencoder는 비지도 학습 모델로, 입력 받은 데이터를 학습해 입력된 데이터에 최대한 가까운 데이터를 생성하는 것이 목표이다. 쉽게 말하면 입력을 예측하는 모델로, 주로 이상 감지, 혹은 노이즈 제거에 사용된다. 이상 감지의 경우 예측한 입력 값과 실제 입력 값의 비교를 통해 이상을 감지하는 방식으로 사용된다.

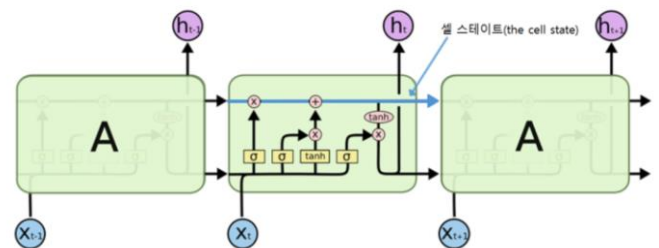


[그림 2] 기본적인 형태의 Autoencoder 구조

Autoencoder의 기본적인 형태는 [그림2]와 같이 입력 받은 데이터를 압축시키는 Encoder와 압축된 데이터를 복원하는 Decoder의 구조로 이루어져 있다. 여기에서 파생된 Convolution Autoencoder, LSTM Autoencoder 등의 모델이 존재한다.

### 2.2.2 LSTM Autoencoder

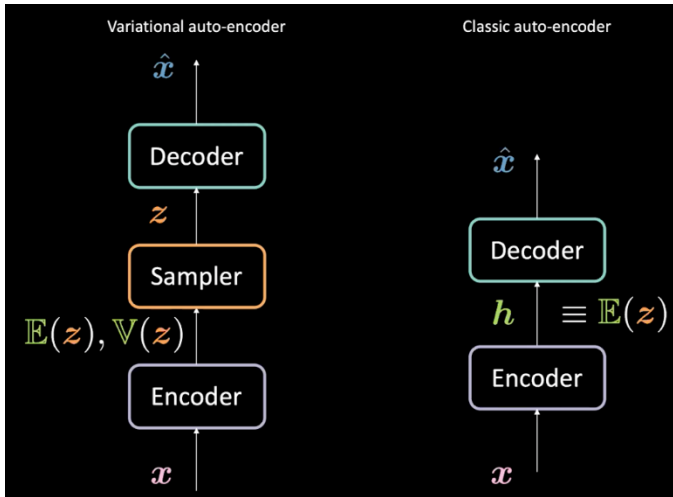
RNN의 경우, 데이터간의 거리가 멀어질수록 데이터의 상관관계가 불명확해지는데 이를 개선한 모델이 바로 LSTM(Long Short Term Memory Network)이다. 가장 기본적인 RNN 모델의 경우, 학습이 거듭될수록 초기의 weight값이 유지되지 않아 학습률이 저하되는 Long-Term Dependency 현상이 생긴다.



[그림 3] LSTM 구조와 셀 스테이트

[그림 3]과 같이 LSTM에서는 Cell State라는 layer를 추가해서 weight를 계속 기억할 것인지 결정하여 이 문제를 개선했다.[2] 이를 통해 직전의 데이터만 참고하는 Vanilla RNN과 달리 LSTM은 cell state에서 다른 weight를 참고할 수 있게 되었다. 이를 이용해 설계한 Autoencoder 모델이 LSTM Autoencoder이다.

### 2.2.3 Variational Autoencoder



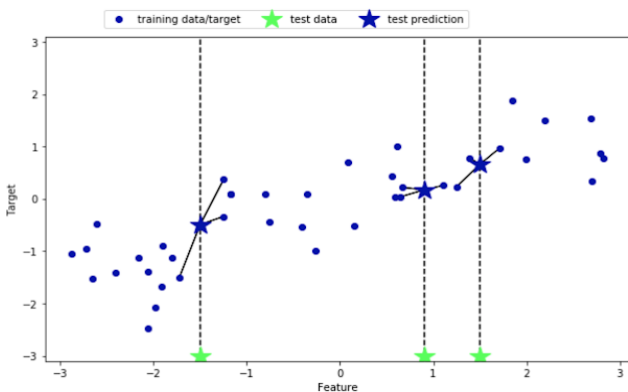
[그림 4] VAE와 기본적인 AE의 비교

Autoencoder의 경우 encoder가 입력  $x$ 에 대한 압축 데이터  $z$ 를 바로 생성하고 그  $z$ 를 가지고 Decoder가 출력  $x'$ 를 생성한다. Variational Autoencoder의 경우에는 바로 생성하지 않고, encoder가 평균 코딩  $\mu$ 와 표준편차  $\sigma$ 를 생성한다. 이후에 가우시안 분포에서 랜덤하게 샘플링 되고, 디코더의 작동방식은 동일하다.

Autoencoder와 목적에서의 차이점은, 입력에 대한 복구가 목표가 아닌, 유사한 새로운 데이터를 생성하는 것이 목표이다.

### 2.2.4 K-NN Regression

K-최근접 이웃(K-Nearest Neighbor, KNN)은 지도 학습 알고리즘 중 하나이다. 어떤 데이터가 주어지면 그 주변의 데이터를 살펴본 뒤 더 많은 데이터가 포함되어 있는 범주로 분류하는 방식이다. K-NN의 특징은 훈련이 따로 필요 없다는 것이다. 훈련 데이터를 기반으로 모델을 만들고 테스트 데이터로 테스트를 하는 방식이다. K-NN은 분류가 아니라 회귀로도 사용할 수 있다.



[그림 5] K-NN Regression을 이용한 예측

## 3. 설명 가능한 딥 러닝을 이용한 심전도 데이터 분석

### 3.1. 비정상 ECG 데이터 생성

정상적인 데이터는 충분하지만, 비정상적인 데이터의 경우 증상도 많고 각 증상에 대한 데이터도 부족해 데이터의 불균형이 우려된다. 기본적인 이상 탐지에는 무리가 없지만, 증상 판별을 위해 실험적으로 Variational Autoencoder를 이용해 이상 데이터에 대한 증식을 시도한다.

### 3.2. 이상 탐지 모델

ECG 데이터를 다양한 종류의 Autoencoder에 입력시켜 이상 탐지 모델을 생성한다. 입력 전의 순수한 데이터를  $x$ , decoder를 통해 출력된 데이터를  $x'$ 라고 했을 경우, 이 둘의 차이를 error라고 칭하며, 정상일 경우 error의 크기가 작고, 비정상일 경우 크게 나오는 점을 이용해 이상 탐지를 하게 된다. 어느 수준의 error까지 정상으로 취급할 것인지를 정하는 threshold를 적절하게 선정하는 것이 중요하다.

### 3.3. 딥러닝 모델 설명 방법

기본적으로는 Autoencoder가 정상분포로 복원한 값과 입력 값을 이미지로 출력해 error를 눈으로 확인 할 수 있도록 하고, 갖고 있는 정상 데이터의 분포와 얼마나 벗어나 있는지를 분포도를 통해 보여준다.

추가적으로, 최근접 이웃 알고리즘을 통해, 갖고 있는 abnormal template에서 어느 case에 가장 가까운지 표현해 증상 판단을 용이하게 한다.

## 4. 결론 및 향후연구

본 논문에서는 설명가능한 딥 러닝 모델을 이용해 심전도 데이터의 분석을 진행해, 사용자가 쉽게 건강의 이상유무를 판별하고, 딥 러닝 모델 신뢰성 저하의 특징인 블랙박스를 해소한다. 딥러닝 모델의 정확도가 높은 수준까지 올라왔지만, 100%라고는 할 수 없으며, 오류는 존재할 수 있다. 설명가능한 모델은 딥 러닝 모델의 판단 과정을 사용자가 검증할 수 있게 하고, 오류에 대응 가능케 한다. 사용자가 쉽게 진단을 이해할 수 있기

때문에, 의료시설이 미흡한 곳에서도 심전도의 측정과 분석이 용이해지고, 보편적 인간의 건강 증진으로 이어질 수 있다.

다만, 발생할 수 있는 abnormal case는 많은 반면에 해당 case들의 데이터는 부족하기 때문에, 현재 심전도 데이터만으로는 진단할 수 있는 증상에 한계가 있고, 가능하더라도 정확도의 보장이 어렵다. 혈액 검사 결과나 기타 진단검사 데이터를 포함한 종합 모델을 사용한다면 좋은 결과를 기대해 볼 수 있을 것이다. 하지만 이를 위해서는 필요로 하는 데이터가 종합적으로 존재하는 환자와 정상인들의 데이터베이스가 확립되어야 한다는 한계점이 있다.

### 참고문헌

- [1] Zhou, B., Liu, S., Hooi, B., Cheng, X. and Ye, J., 2019, August. BeatGAN: Anomalous Rhythm Detection using Adversarially Generated Time Series. In *IJCAI* (pp. 4433-4439)
- [2] Jeyakumar, J.V., Noor, J., Cheng, Y.H., Garcia, L. and Srivastava, M., 2020. How can i explain this to you? an empirical study of deep neural network explanation methods.
- [3] Kuznetsov, V.V., Moskalenko, V.A., Gribanov, D.V. and Zolotykh, N.Y., 2021. Interpretable Feature Generation in ECG Using a Variational Autoencoder. *Frontiers in Genetics*, 12.
- [4] Golany, T., Radinsky, K. and Freedman, D., 2020, November. Simgans: Simulator-based generative adversarial networks for ecg synthesis to improve deep ECG classification. In *International Conference on Machine Learning* (pp. 3597-3606). PMLR.
- [5] Ribeiro, A.H., Ribeiro, M.H., Paixão, G.M.M. *et al.* Automatic diagnosis of the 12-lead ECG using a deep neural network. *Nat Commun* 11, 1760 (2020).