

IoT 융합실험 REPORT

ECG 생체정보를 활용한 AutoEncoder Based Anomaly Detection

전자공학부 지능IoT 20184234 이시현

What is ECG?

- ECG : 심장의 전기적 활동을 측정하고 기록한 결과물로, 심장 상태를 평가하고 심전도 패턴을 분석
- ECG를 분석하여 심장 박동의 정규성, 이상 심전도 패턴, 부정맥, 심방세동 등과 같은 심장 질환을 평가할 수 있음.
- ECG는 운동 부하 테스트, 심장 수술 전후 모니터링, 심장약물의 효과 평가 등에도 사용
- ECG를 활용한 사례로는 아래와 같은 것들이 있음

이상탐지

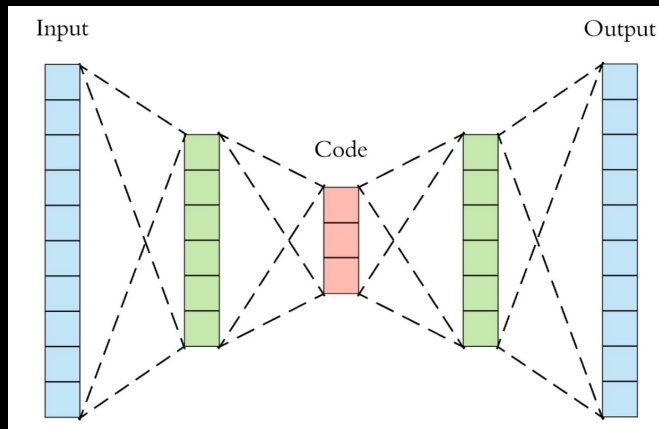
심장질환진단

실시간 모니터링

개인 맞춤 치료 및 예방

What is AutoEncoder?

- AutoEncoder : 딥러닝 기반의 비지도학습 모델로, 입력 데이터를 압축하여 재구성
- Encoder와 Decoder로 구성되어 있으며, Encoder에서는 데이터를 압축하고, Decoder는 저차원 데이터를 원래의 표현으로 복원



- 이상 ECG 신호는 정상적인 패턴과 다른 특징을 가지고 있기 때문에, AutoEncoder는 이상치를 탐지하고 재구성 오차를 측정하여 이상탐지 수행

프로젝트 선정 배경

지난 4월 이상탐지를 사용해 방사선 계측 분야에 도움이 되고자 한차례의 논문을 작성하였음. 그러나 평소 선량률 예측에 대한 프로젝트만 진행했어서 다양한 분야에 이상탐지를 접할 기회가 많았음.

ECG 데이터를 사용해 딥러닝 모델을 사용하여 무엇인가를 예측하기에 앞서 이상 탐지는 성능 개선을 위해 꼭 필요하다고 생각. 이에 연구 경험을 늘리고자 ECG 데이터셋을 활용해 연구 분야인 이상탐지의 프로젝트를 해보고자 하였음.

One-Class SVM Based Anomaly Detection for LSTM-Based Radiation Prediction

Si Hyun LEE

Chosun University, [Gwangju](#), Korea

This study emphasizes the importance of identifying pollutants and their effects on our environment through environmental monitoring. Environmental monitoring aims to continuously monitor various environmental factors such as air, water quality, and radiation, and to prevent and solve environmental problems by collecting and analyzing information on environmental changes.

In this study, we confirmed that a multivariate data model including both radiation dose rate and weather data is required for radiation dose rate prediction. To this end, LSTM was used for seasonal data prediction, and the performance of the model was improved by removing abnormal values of dose rate using one-class SVM.

The One-Class SVM learns using only one class data, which sets the dose rate to one class and finds the hyperplane. The hyperplane aims to maximize the area where dose rate data is located and to maximize the margin with other data.

Anomaly detection model learns with dose rate data only and models the distribution of dose rate data using one-class SVM. This model is used to classify new data, and data located on one side of the hyperplane is determined to be outliers. Therefore, by detecting and removing outlier data for dose rates, the performance of the radiation dose prediction model can be improved.

After detecting and removing outliers using the One-Class SVM, the radiation dose rate was predicted using the LSTM. A model that uses input data and previous states to predict the next state. You can consider the temporal dependence of the input data by remembering the information that occurred in the previous state and considering it in the next state. This was to predict seasonal radiation dose rate data.

Finally, the performance of the radiation dose rate prediction model using one-class SVM and LSTM was evaluated. The performance of the One-Class SVM and LSTM models was evaluated through various metrics such as MSE and confusion matrix, accuracy, precision, reproduction rate, and F1 score, and the performance of the One-Class SVM was 89.87%, accuracy of 1.0, reproduction rate of 0.8987, and F1 score of 0.9467. The MSE of the LSTM was 0.00156, which showed a lower value than that of the model without the One-Class SVM. Therefore, One-Class SVM performed well in removing outliers from radiation dose rate data through outlier detection, and LSTM performed well in predicting seasonal radiation dose rate data considering time dependence. Therefore, this study demonstrated that the radiation dose rate prediction model combining one-class SVM and LSTM has high performance.

Through this study, we developed a dose rate prediction model that combines one-class SVM

Process

데이터 수집

데이터 변환

이상탐지 모델 학습

모델 평가



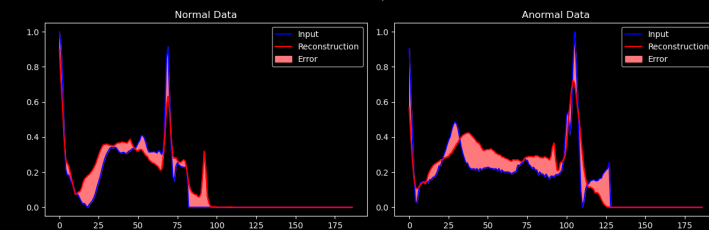
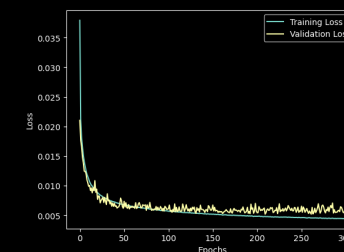
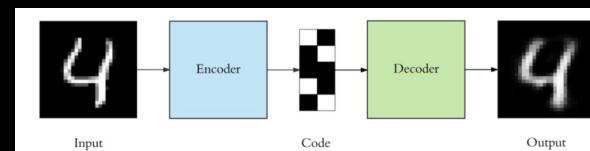
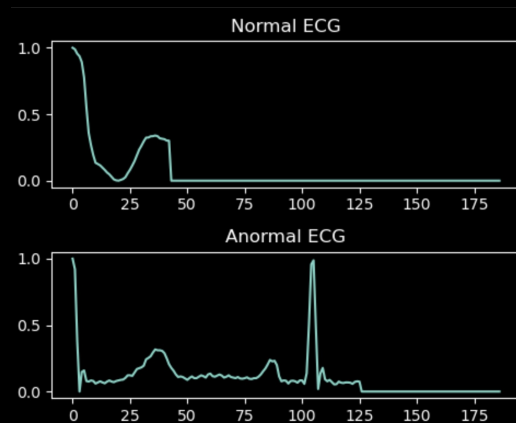
Dataset

ECG dataset

by Devavrata Tripathy

2 years ago • 3 MB • 32

[ECG dataset](#)



- Kaggle 사이트 이용

- Min-Max Regularization
- Normal vs Anormal 분리

- AutoEncoder 사용

- Train loss / val loss
- Find Anormal Data

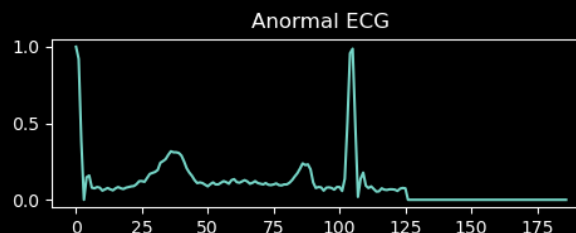
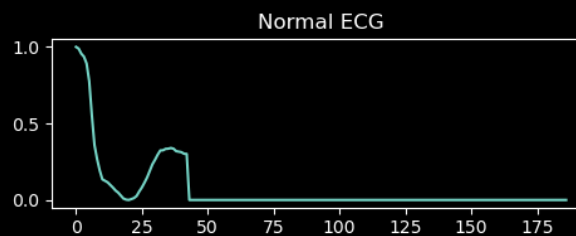
데이터 수집 및 변환

- Kaggle의 ECG 데이터셋 사용(<https://www.kaggle.com/datasets/devavratatripathy/ecg-dataset/code?datasetId=1103711&sortBy=voteCount>)
- 학습 vs 테스트 데이터 분리 (87554, 188) → (70043, 187)
(17511, 187)
- Min-Max 정규화 수행 : AutoEncoder는 입력 데이터와 출력 데이터가 동일한 형태를 가져야 하므로, 입력 데이터의 범위를 조절해야 함. 또한 이상치의 영향을 줄이기 위함. 0.0
1.0
- Normal vs Anormal 데이터 분리 : Train과 Test 데이터에서 정상값과 이상값에 해당하는 값으로 분리. 이후 정상값인 Normal Train 데이터만 사용하여 AutoEncoder를 학습하는 데 사용

```
print(normal_train.shape, normal_test.shape)  
print(anormal_train.shape, anormal_test.shape)
```

✓ 0.0s

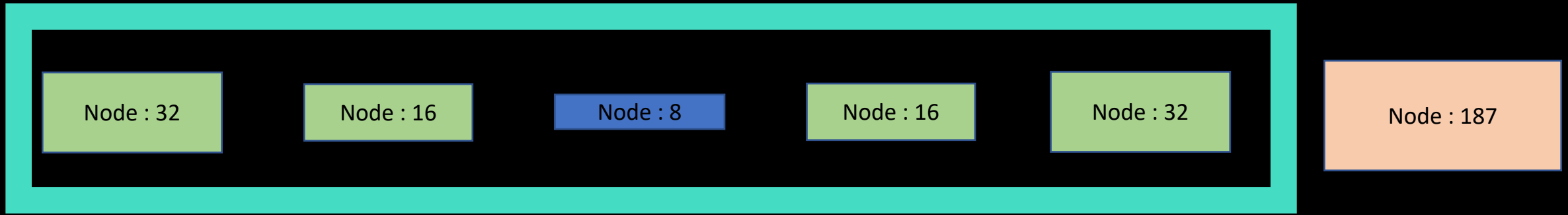
```
(12026, 187) (3057, 187)  
(58017, 187) (14454, 187)
```



모델 학습_개선 전

Input

Output



Epoch : 500

Batch Size : 100

Optimizer : Adam

모델 학습_개선 후

Input

Dropout(0.2)

Dropout(0.2)

Output

Node : 128

Node : 64

Node : 32

Node : 16

Node : 8

Node : 16

Node : 32

Node : 64

Node : 128

Node : 187

Epoch : 500

Batch Size : 100

Optimizer : RMSprop

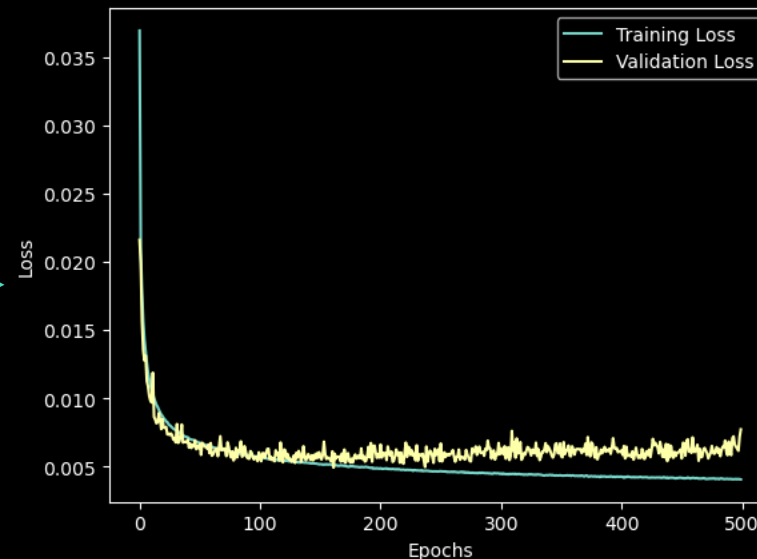
모델 평가 및 결과 설명

개선 전

Train Loss : 0.0301
Val Loss : 0.016

개선 후

Train Loss : 0.004
Val Loss : 0.007



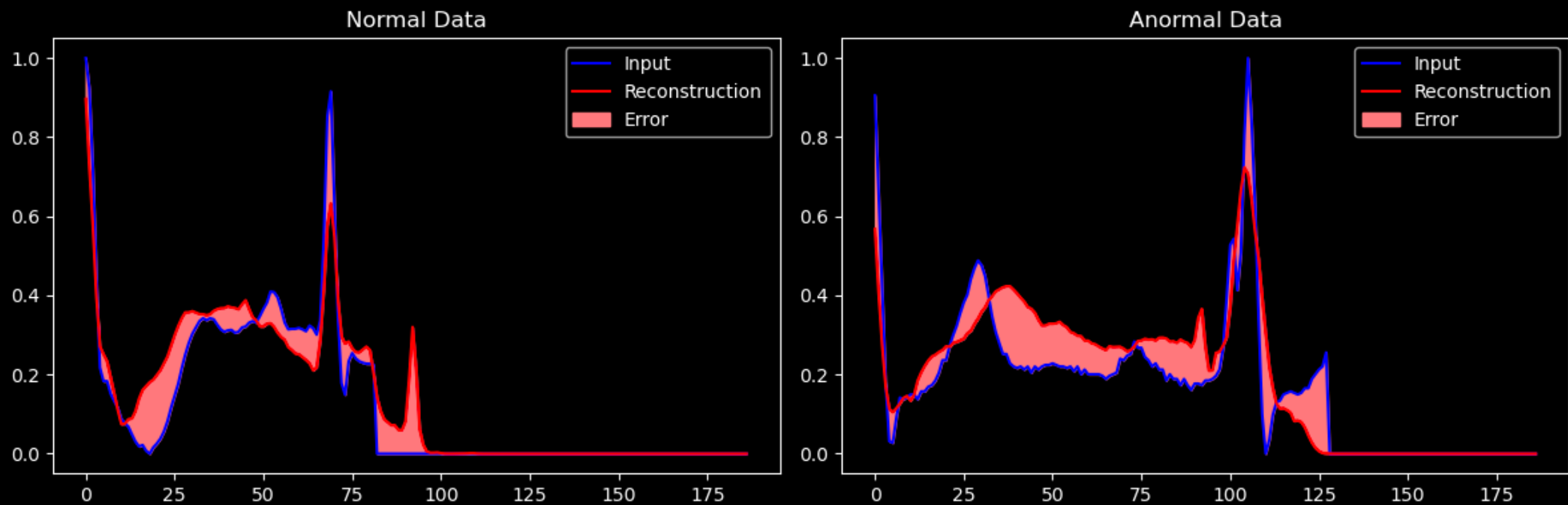
변동사항 설명

개선 전에는 Train Loss는 줄어들이지만 Val Loss는 줄지 않는 Overfitting 문제가 발생하였음. 따라서 Loss 값을 0에 수렴하게 만들고자 **Dropout**을 적용하였고, 모델의 복잡성을 증가시키기 위해 **노드 수를 증가**시켰다. 또한 Adam Optimizer 대신 **RMSprop**을 사용하였다. 이외에도 Batch Size 크기를 바꾸거나 다른 정규화를 써보는 등을 시도했지만 성능 변화에는 큰 차이가 없어 이 모델에서는 사용하지 않기로 결정함.

결과 설명

Train / Val Loss 모두 0에 수렴

이상 탐지 결과



AutoEncoder 모델로 ECG 데이터의 Input 데이터(원본), AutoEncoder가 재구성한 Reconstruction 데이터의 그래프를 그린 결과이다. AutoEncoder에서 이상치를 제외한 부분을 출력하기에 Error가 발생하는 것을 확인할 수 있다.

프로젝트 소견

1. 프로젝트 목표:

- 심전도 신호에서 이상 탐지 및 심장 질환 진단을 자동화하는 AI 모델 개발

2. 프로젝트 주요 내용:

- ECG 데이터 전처리: 심전도 신호를 정규화하고, 잡음 제거 및 데이터 정제 작업 수행
- AutoEncoder 모델 학습: 비지도학습 기법을 사용하여 정상적인 심전도 패턴을 학습하고, 이상 탐지에 활용
- 모델 평가 및 성능 개선: 모델의 성능을 확인하고, 성능 개선을 위한 실험 및 수정 작업 수행

3. 프로젝트의 한계 및 해결 방안

- 데이터 양과 다양성 부족: 추가적인 데이터 수집 및 다양한 심장 질환 케이스 포함
- 모델 성능 향상: 적절한 하이퍼 파라미터 선정

4. 소감

- 적절한 하이퍼파라미터 선정에 어려움을 느꼈음. 평소 알고있던 이론을 실무에 적용하여 사용함으로써 각 도메인에 맞는 모델 선정과 하이퍼파라미터 선정이 중요하다는 것을 기존 연구 도메인과 다른 데이터를 사용함으로써 느끼게 됨