대본\_한글

먼저 Keras를 이용한 anomaly detection의 예제를 연습해보았다.

Numenta Anomaly Benchmark(NAB) dataset을 사용했다. It provides artifical timeseries data containing labeled anomalous periods of behavior. Data are ordered, timestamped, single-valued metrics.

===========================

Timeseries란?

시간은 시계열 패턴이다. 시계열 패턴은 연속적이다. 연속적이다 라는 말은 선후 관계가 있다는 말이다. 선후 관계를 통해 추정과 조합이 가능하다.

AutoEncoder란?

오토인코더의 경우 보통 이미지의 생성이나 복원에 많이 사용되며 이러한 구조를 이어받아 대표적인 딥러닝 생성 모델인 GAN으로까지 이어졌는데 이러한 자기 학습 모델은 Anomaly Detection 분야에서도 널리 사용되고 있다.

대표적으로 이미지 분야에서도 정상적인 이미지로 모델 학습 후 비정상적인 이미지를 넣어 이를 디코딩 하게 되면 정상 이미지 특성과 디코딩 된 이미지 간의 차이인 재구성 손실(Reconstruction Error)를 계산하게 되는데 이 재구성 손실이 낮은 부분은 정상, 재구성 손실이 높은 부분은 이상으로 판단할 수 있다.

이러한 Anomaly Detection은 이미지뿐만 아니라 이제부터 살펴보고자 하는 시계열 데이터에도 적용이 가능하다. 예를 들어 특정 설비의 센서를 통해 비정상 신호를 탐지하고자 한다면 Autoencoder를 LSTM 레이어로 구성한다면 이러한 시퀀스 학습이 가능하게 된다.

이를 통해 정상 신호만을 이용하여 모델을 학습시켜 추후 비정상 신호가 모델에 입력되면 높은 reconstruction error를 나타낼 것이므로 이를 비정상 신호로 판단할 수 있게 된다.

LSTM Autoencoder는 시퀀스 데이터에 Encoder-Decoder LSTM 아키텍처를 적용하여 구현한 오토 인코더이다. 모델에 입력 시퀀스가 순차적으로 들어오게 되고, 마지막 입력 시퀀스가 들어온 후 디코더는 입력 시퀀스를 재 생성하거나 혹은 목표 시퀀스에 대한 예측을 출력한다.

위에서 설명한 것과 마찬가지로 LSTM Autoencoder 학습 시에는 정상 신호의 데이터로만 모델을 학습시키게 된다. encoder와 decoder는 학습이 진행될 수 록 정상신호를 더 정상 신호 답게 표현하는 방법을 학습하게 될 것이며 최종적으로 재구성한 결과도 정상 신호와 매우 유사한 분포를 가지는 데이터일 것이다. 그렇기 때문에 이 모델에 비정상 신호를 입력으로 넣게 되면 정상 분포와 다른 특성의 분포를 나타낼 것이기 때문에 높은 reconstruction error를 보이게 될 것이다.

코드 리뷰?

가장 처음엔 라이브러리를 설치한다

그 후 데이터를 로드해서 데이터의 패턴들과 데이터의 anomaly가 무엇인지 확인했다.

모델을 학습시킬 만할 학습데이터를 준비하면서, 시퀀스데이터를 확인함

모델을 구성하고 학습시킴

anomaly detect를 수행하고

재구성된 것과 비교함

최종적으로 plot을 찍어볼 것이다.

p.1

Hello. My name is 김영태, and I’m going to start the presentation now.

p.4

(페이지 두개 건너뛰고)

Prior to the actual implementation of the project, we found and practiced examples for anomaly detection. What we did was anomaly detection example using Keras. Through this example, we were able to understand how the abnormal detection system works overall

We practice example of anomaly detection in convolutional autoencoder model.

We use Numenta Anomaly Benchmark(NAB) dataset. It provides artifical timeseries data containing labeled anomalous periods of behavior. Data are ordered, timestamped, single-valued metrics.

First, We installed the library that we needed for example. And we loaded the Numenta Anomaly dataset to check the patterns of the data and what the anomaly of the data was. There are two time series data we received. One of them is data that has normal properties. the other one is data that has abnormal properties.

p.5

We created a model for learning and put data into this model to learn.

And by outputting the plot, We could see training loss and validation loss.

We will detect anomalies by determining how well our model can reconstruct the input data.

First, Find MAE loss on training samples.

and Second, Find max MAE loss value. This is the worst our model has performed trying to reconstruct a sample. We will make this the threshold for anomaly detection.

and finally, If the reconstruction loss for a sample is greater than this threshold value then we can infer that the model is seeing a pattern that it isn't familiar with. We will label this sample as an anomaly.

p.6

Starting next week, we will look for various ways to detect abnormal signal in time series data and review each auto encoder model one by one.

가장 처음엔 라이브러리를 설치한다

그 후 데이터를 로드해서 데이터의 패턴들과 데이터의 anomaly가 무엇인지 확인했다.

모델을 학습시킬 만할 학습데이터를 준비하면서, 시퀀스데이터를 확인함

모델을 구성하고 학습시킴

anomaly detect를 수행하고

재구성된 것과 비교함

최종적으로 plot을 찍어볼 것이다.