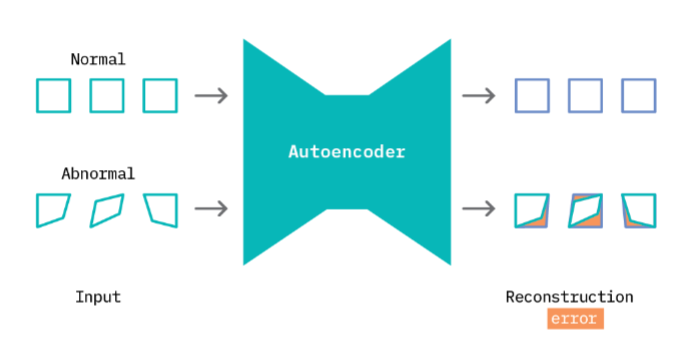
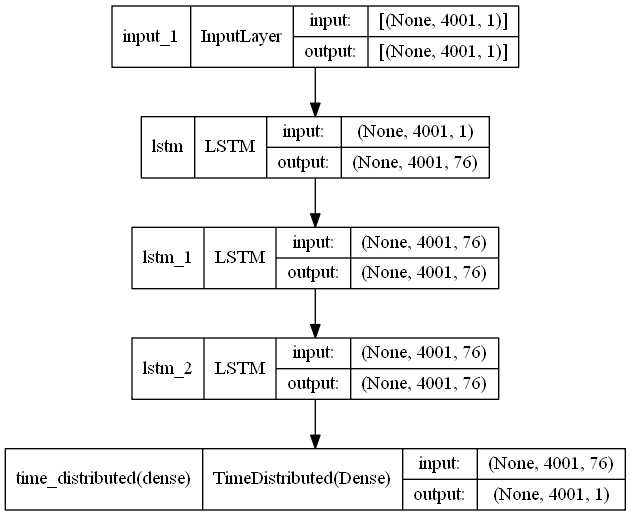
Anomaly detection – Reconstruction error



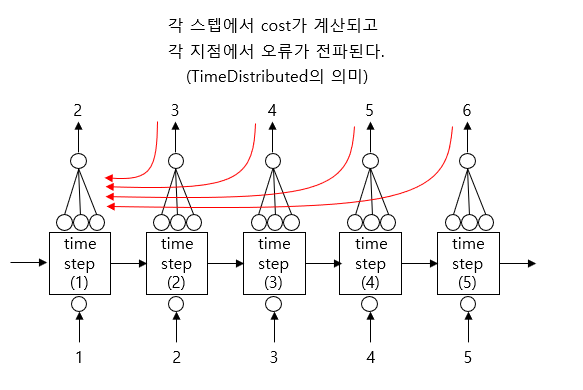
Anomaly detection 모델은 Normal data들을 다시 복원하는 방법을 학습한다. 이렇게 되면 Abnormal data가 들어왔을 때 input data와 reconstruction data 사이의 차이가 커지게 되고 이를 anomaly score로 사용할 수 있다.

LSTM model

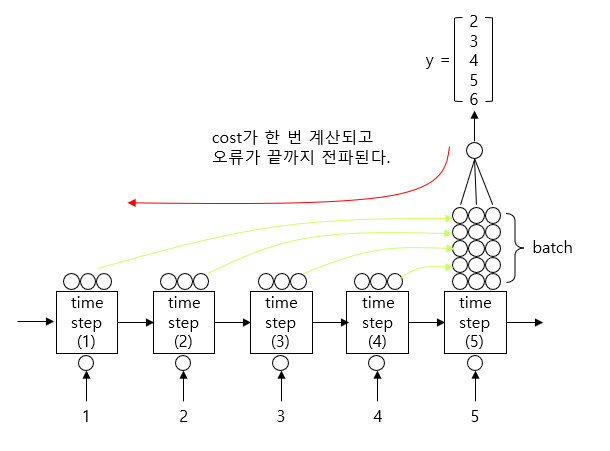


Time distributed layer의 필요성

* Time distributed layer가 있을 때

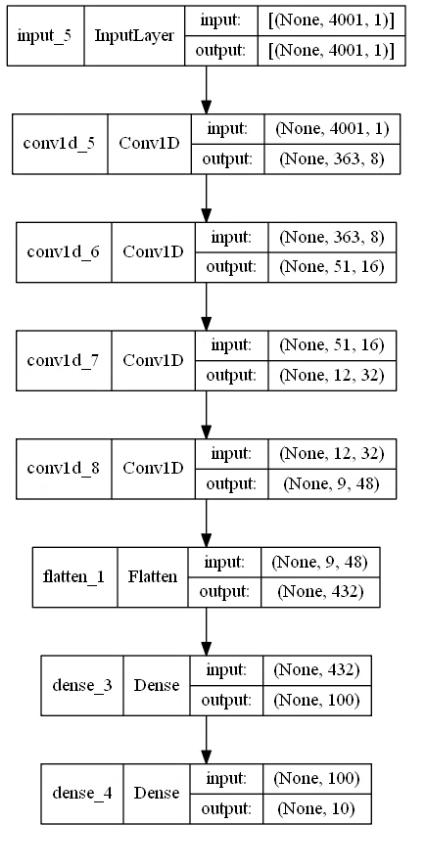


- Time distributed layer가 없을 때



VAE(LSTM or Conv) model

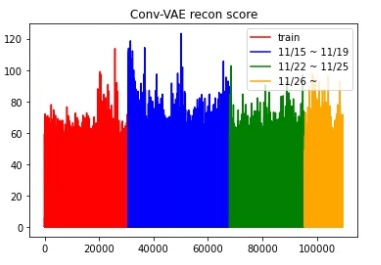
* Encoder

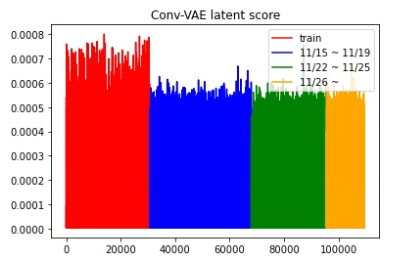


* Sampling
* Decoder
* 테이블이(가) 표시된 사진

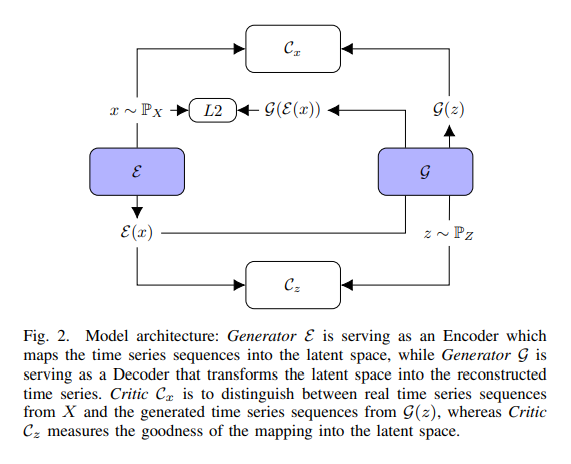
  자동 생성된 설명

아래와 같이 reconstruction error와 latent error 두 가지 방식으로 anomaly score를 계산할 수 있음





TadGAN model



Encoder: time series sequence를 latent space로 mapping

Generator(decoder): latent space의 data를 time series로 reconstruct

Critic X: 실제 input data와 reconstruct data를 구분

Critic Z: encoder를 통과한 input data와 z~Pz에서 생성된 data를 구분

이상탐지 방법: critic X score와 reconstruct score를 결합해서 사용

성능비교

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | LSTM | LSTM-VAE | CONV-VAE | CONV-AE | TadGAN |
| SMD | 0.77 | 0.73 | 0.71 | 0.64 | 0.6 |
| SMAP | 0.57 | 0.68 | 0.52 | 0.53 | 0.59 |
| MSL | 0.55 | 0.52 | 0.62 | 0.6 | 0.56 |
| SWaT | 0.81 | 0.72 | 0.74 | 0.69 | 0.81 |
| 평균 | 0.675 | 0.6625 | 0.6475 | 0.615 | 0.64 |

\* TadGAN의 경우 단순히 모델의 loss function 값이 아닌 critic과 generator의 loss function을 비교하면서 학습을 진행해야 되는데 시간 상 현재는 그러한 작업들을 진행하지 못한 상태. 때문에 위의 값에서 개선될 여지가 많음.

\* VAE 모델 또한 논문과 같은 방식으로 구현하지 않고 단순화시켜서 구현했기 때문에 개선될 여지가 있음.

\* Transformer 모델의 경우 개발 완료 후 ROC-AUC score는 계산하지 않았었는데 추후 실험 진행 후 점수를 비교해볼 예정