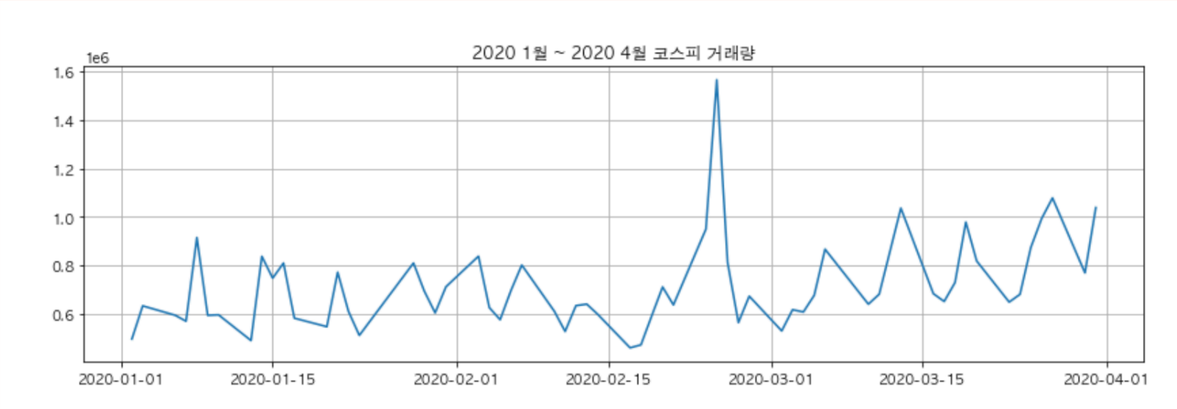
비지도학습
시계열 이상치 탐지

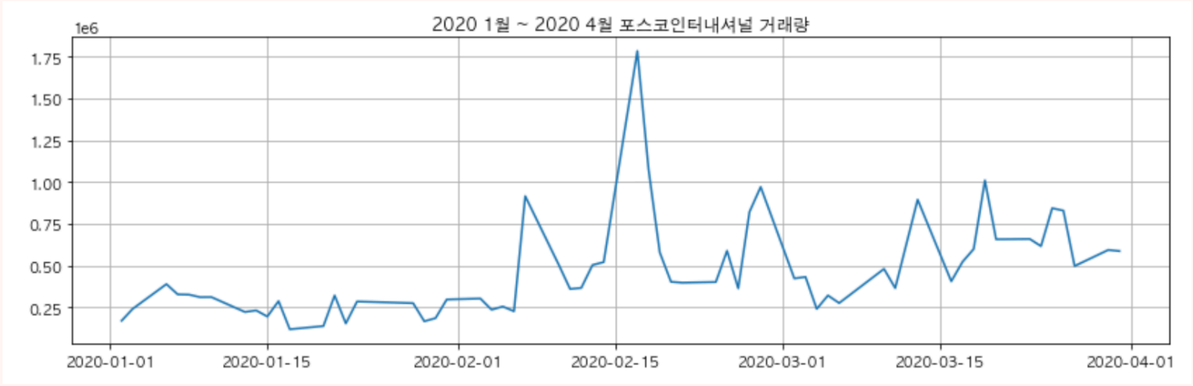
박현우

INPUT 데이터

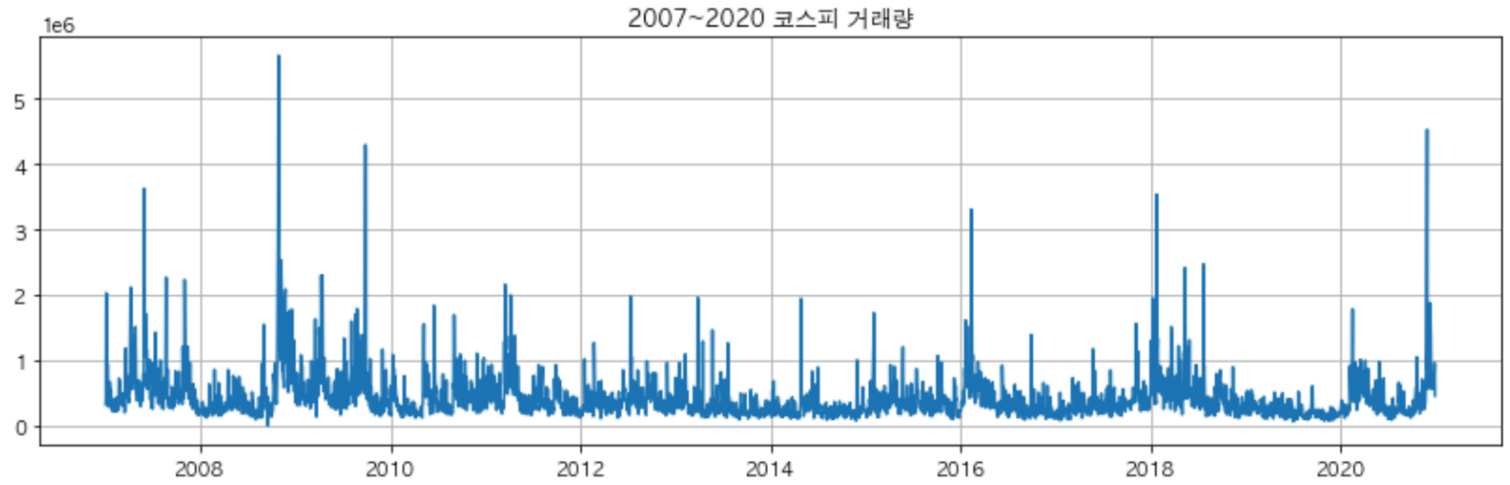
1. 일별 코스피 거래량 데이터
2. 일별 코스피 200 종목의 거래 데이터 (거래량, 일별 수익률 , 고가와 저가의 차이 등)
3. 이상치 모델링을 위한 위의 데이터들의 특징
    - 1개의 변수의 이상탐지가 아닌 다변량 변수를 사용함
    - 시계열 특성을 지님 -> 시계열 이상치 탐지방법
    - 이상치인지 아닌지 labeling이 되어있지 않음 -> Unsupervised Learning 이나 Semi-supervised learning 방법을 사용해 이상치 탐지

코스피 거래량과 종목별 거래량의 상관성





코스피 거래량과 종목별 거래량의 상관성



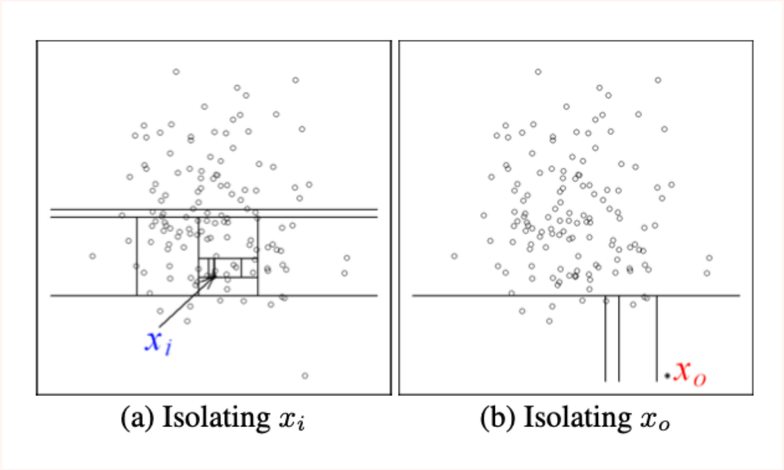
- 코스피 거래량과 포스코인터내셔널의 거래량 상관계수 : **0.173**
- 코스피 일별 수익률과 포스코인터내셔널 일별 수익률 상관계수 : **0.518**
- 코스피 거래량과 코스피 **870**개 종목별 거래량의 상관계수 평균값 : **0.136**
- 코스피 일일 수익률과 코스피 **870**개 종목별 일일 수익률의 상관계수 평균값 : **0.334**

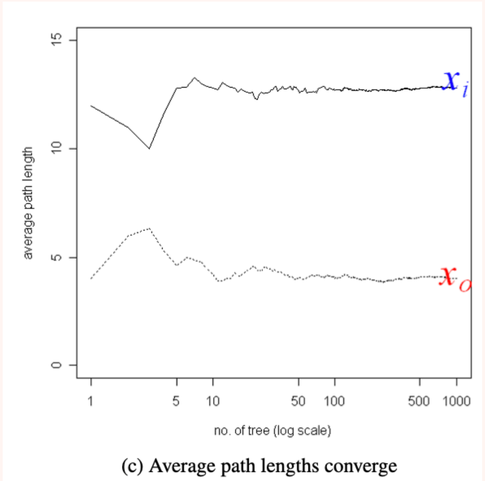
비지도학습 이상치 탐지

1. **Density-estimation method**
    - **LOF , COF -** 지역 밀도와 연결성을 계산하여 이상치 탐지
    - **DAGMM -** 밀도 추정을위해 **Gaussian Mixture Model**을 사용
2. **Model Based Method**
    - **Isolation Forest -** 결정트리와 **Bagging**을 사용해 이상치 탐지
3. **Clustering Based method**
    - **SVDD -** 클러스터 중심으로부터 거리를 공식화하여 **Anomaly score** 계산을 통해 이상치탐지
    - **Deep SVDD -** 정상 데이터만을 학습하여 정상 **feature**를 둘러싸는 최적의 구를 찾아내는것 **(Semi Supervised learning)**
4. **Reconstruction-based model**
    - **LSTM-VAE** 모델 **: LSTM**기반으로한 일시적인 모델링과 **VAE**를 활용해 **reconstruction error**을 계산하여 이상치 탐지
        - **Reconstruction error**는 **anomaly score**로써 임계값을 기준으로 이상 탐지
5. **Association based model**
    - **Anomaly Transformer : Time Series Anomaly Detection With Association Discrepancy (2022)** 논문에서 설명된 모델

ISOLATION FOREST

- 이상치의 특징 중 소수의 데이터 , 정상데이터와 다른 속성 값을 지닌다는 특징을 이용하여 모델링
- **Isolation Forest**는 임의의 변수를 선택해 임의의 값으로 **Isolation Tree**를 **split**하여 데이터를 “고립” 시키는 방식을 사용함 -> 즉 고립된 데이터는 이상치로 분류를 시킨다.
- **Anomaly score : 충분히 많은 Tree**를 생성해 데이터를 고립시키위해 **split**을 한 횟수의 누적평균을 이용
  - 누적 **split** 평균값이 작다면 **Anomaly score**값은 큰 값을 가짐





Reference :Liu, Fei Tony, Kai Ming Ting, and Zhi-Hua Zhou. ”Isolation forest.” In 2008 Eighth IEEE International Conference on Data Mining, pp. 413-422. IEEE, 2008.

ISOLATION FOREST

Isolation Forest 예외

1. 모든 변수가 같은 값을 지니는 데이터들은 **split** 할 수가 없음 -> **split** 중단

2. 너무 많은 **split**은 비효율적 -> **limit depth**를 **hyperparameter**로 사용하여 **Tree** 깊이에 제한을 둔다

ISOLATION FOREST

Anomaly Score

$$s(x,n) = 2^{-\frac{E(h(x))}{c(n)}}$$

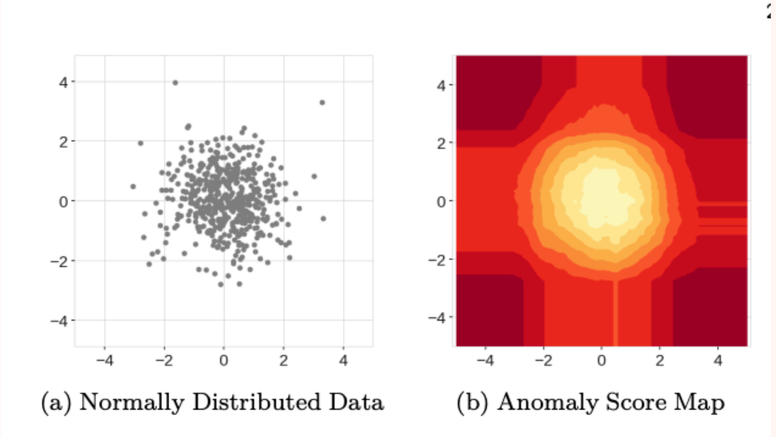
$$c(n) = 2H(n-1) - (2(n-1)/n)$$

$$H(i) = ln(i) + 0.5772156649$$

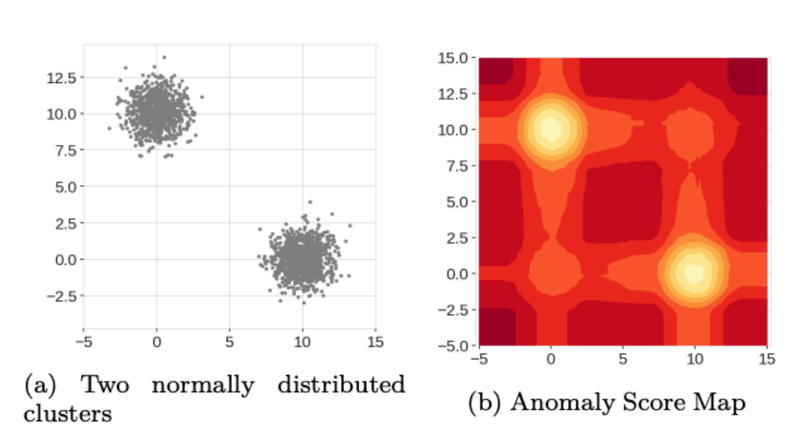
- x : data point , n : number of trees
- c(n) : average path length of unsuccessful search in Binary search Tree (오일러 상수 사용)
- h(x) : isolation tree’s path length to isolate x point

EXTENDED ISOLATION FOREST (2018)

Isolation Forest의 문제점



(a) Normally Distributed Data (b) Anomaly Score Map

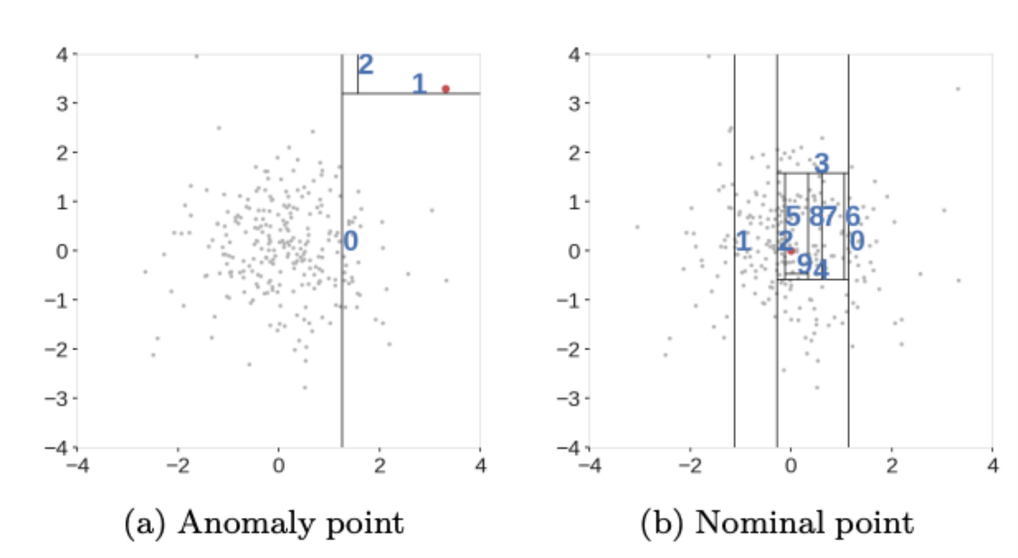


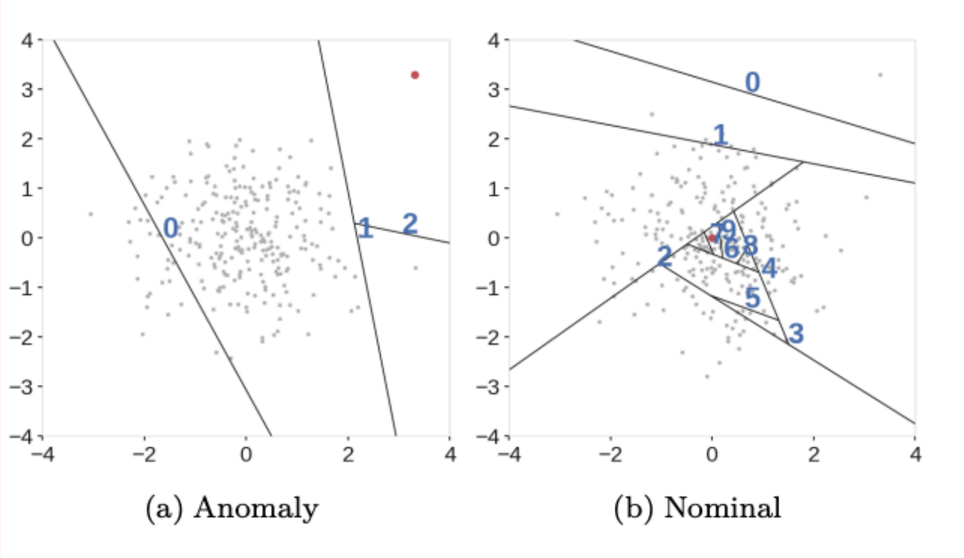
(a) Two normally distributed clusters (b) Anomaly Score Map

Reference : Extended Isolation Forest

EXTENDED ISOLATION FOREST(2018)

개선 방안





Extended Isolation Forest

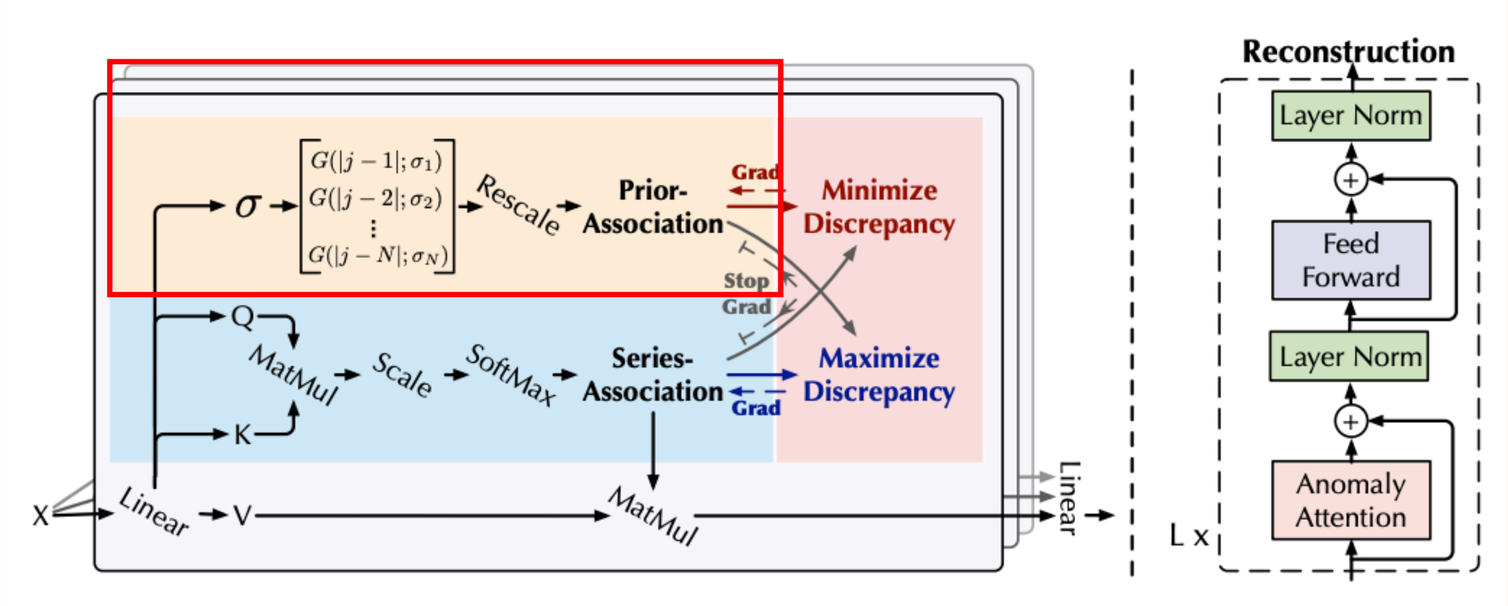
ANOMALY TRANSFORMER

- 기존 비지도학습의 이상탐지의 단점
  - **Pointwise representation**은 복잡한 시점에 정보력이 덜할수 있고 정상 데이터에게 지배당할수 있다
  - **Reconstruction error => point-wise**된 계산을 하여 주변맥락을에 대한 고려를 못한다.
- **Anomaly Transformer**의 목적은 각 시점들의 **Association**을 이용하여 일시적인 맥락과 전체적인 맥락을 모두 고려하기 위한 모델이다.

ANOMALY TRANSFORMER

- **Prior Association : 주변 시점의 attention을 반영한 Association 분포**
- **Series Association : Transformer의 Self-Attention기법을 이용하여 각 시점의 association 분포를 이용해 트렌드에 대한 정보를 묘사한 Association 값분포**
  - **이상치는 드물기 때문에 전반적인 시계열에는 약한 association을 가지고 인접한 영역에 대해서는 강한 association을 가진다.**
  - **Association Discrepancy는 Priori association 과 Series Association간의 거리로 정의함**
- **MiniMax Strategy**
  - **이상치 탐지를 위해 Series Association 은 Minimize , Prior Association은 Maximize 하는 전략을 택함**

ANOMALY ATTENTION



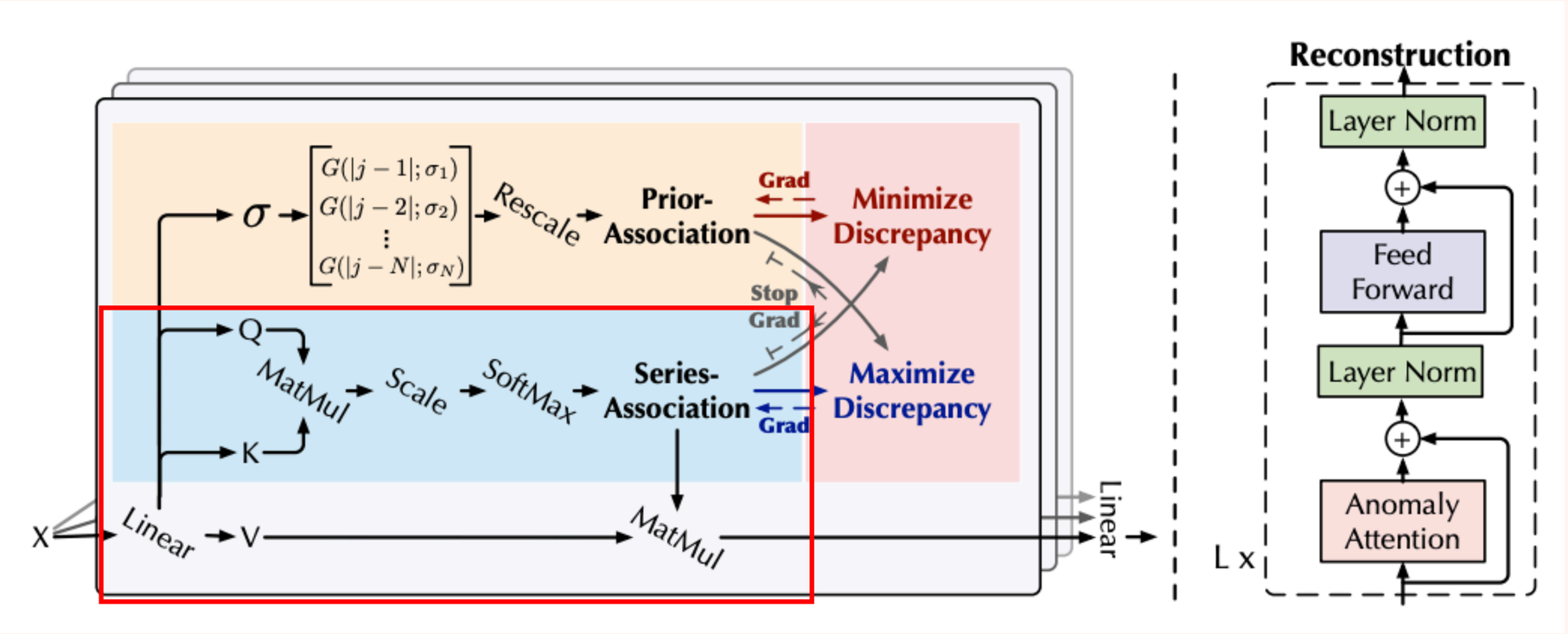
- **Prior Association : 시점간의 거리차이를 이용해 Gaussian Distribution에 적절한 learnable parameter σ 를 학습한다**

- **Rescale : 연속성을 띄는 association weight 분포를 이산형분포로 바꾸기위해 해당 열의 합으로 나눈다**

$$\text{Prior-Association: } \mathcal{P}^l = \text{Rescale}\left(\left[\frac{1}{\sqrt{2\pi}\sigma_i}\exp\left(-\frac{|j-i|^2}{2\sigma_i^2}\right)\right]_{i,j\in\{1,\cdots,N\}}\right)$$

Anomaly Transformer : Time Series Anomaly Detection With Association Discrepancy (2022)

ANOMALY ATTENTION

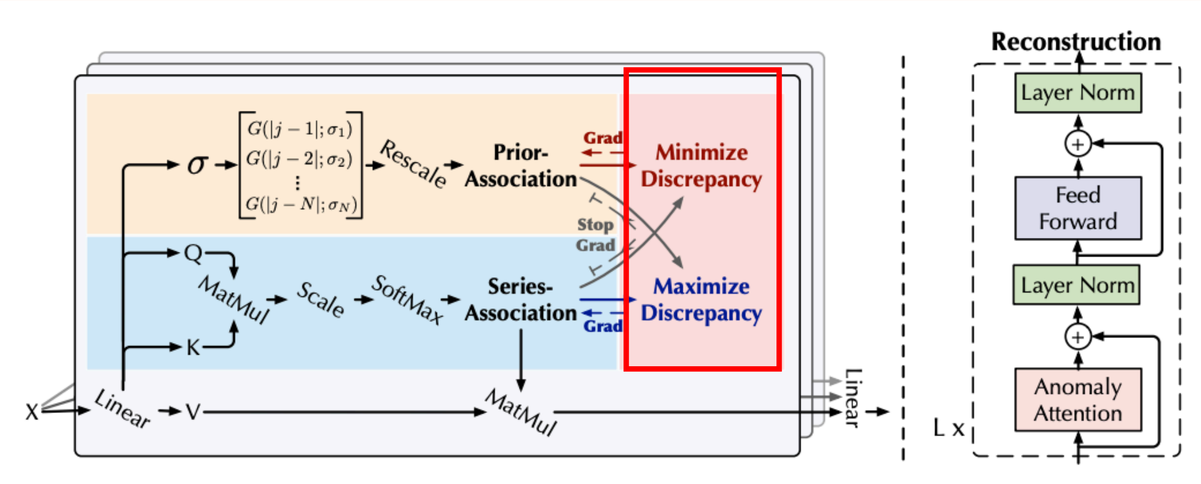


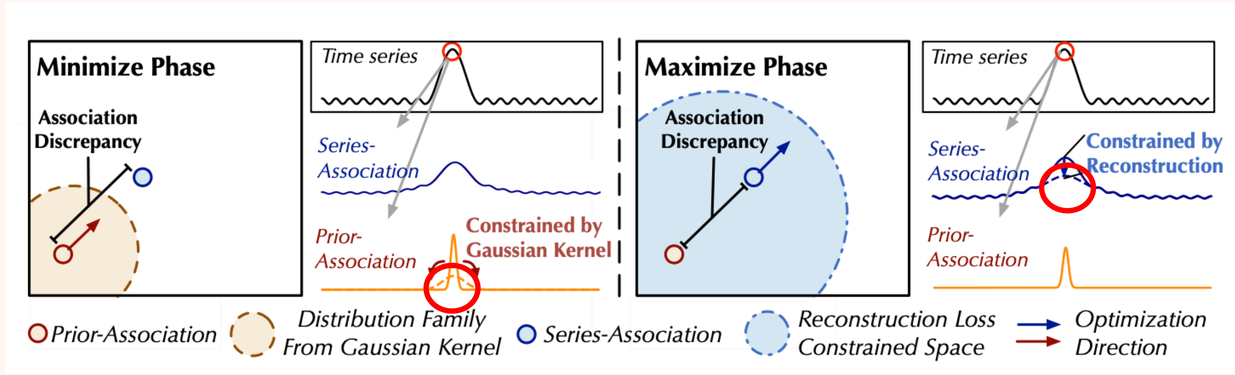
- Series Association은 Transformer을 이용해 전체 raw series로부터 Association을 학습한다

$$\text{Series-Association: } \mathcal{S}^{l} = \text{Softmax}\left(\frac{\mathcal{Q}\mathcal{K}^{\mathrm{T}}}{\sqrt{d_{\text{model}}}}\right)$$

Anomaly Transformer : Time Series Anomaly Detection With Association Discrepancy (2022)

ANOMALY ATTENTION





- **Minimize phase : Prior Association을 Series Association에 가깝게 하여 Association Discrepancy 를 최소화 하는 방향**

- **Maximize phase : Series Association을 Prior Association에서 멀어지게 하여 Association Discrepancy를 최대화 하는 방향 => Series Association이 인접지역이아닌 더 넓은 범위에 Attention을 주도록 함**

Anomaly Transformer : Time Series Anomaly Detection With Association Discrepancy (2022)

ASSOCIATION DISCREPANCY

$$\mathrm{AssDis}(\mathcal{P},\mathcal{S};\mathcal{X}) = \frac{1}{L}\sum_{l=1}^{L}\left(\mathrm{KL}(\mathcal{P}^l \| \mathcal{S}^l) + \mathrm{KL}(\mathcal{S}^l \| \mathcal{P}^l)\right)$$

$$D_{KL}(P||Q) = \sum_i P(i) log \frac{P(i)}{Q(i)}$$

$$= \sum_i P(i) log P(i) - \sum_i P(i) log Q(i)$$

$$= -H(P) + H(P,Q)$$

- S : Series Association , P : Prior Association
- KL : KL divergence (정보엔트로피와 크로스엔트로피로 분해)
  - 두 association 간의 유사도를 측정하기 위한 방법

추가 학습

- **Anomaly Transformer의 알고리즘 이해**
  - **Transformer에대한 자세한 이해를 위해 “Attention is all you need” 논문 분석**
  - **Attention Mechanism에 대한 이해 필요**
  - **“Transformer in Time Series : A Survey” 논문 분석**

REFERENCE

1. Liu, Fei Tony, Kai Ming Ting, and Zhi-Hua Zhou. ”Isolation forest.” In 2008 Eighth IEEE International Conference on Data Mining, pp. 413-422. IEEE, 2008.

2. Extended Isolation Forest

3. Attention Is All You Need (2017)

(Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Lukasz Kaiser, Illia Polosukhin)

3. Anomaly Transformer : Time Series Anomaly Detection With Association Discrepancy (2022)