

---

# Multivariate Time-series Anomaly Detection via Graph Attention Network

---

School of Industrial and Management Engineering, Korea University

Hansam Cho

# Contents

---

- ❖ Research Purpose
- ❖ Time-series Anomaly Detection
- ❖ Multivariate Time-series Anomaly Detection via Graph Attention Network(MTAD-GAT)
- ❖ Experiments

# Research Purpose

- ❖ Multivariate Time-series Anomaly Detection via Graph Attention Network(2020, ICDM)
  - Microsoft 소속의 저자, 2022년 04월 11일 기준으로 44회 인용

## Multivariate Time-series Anomaly Detection via Graph Attention Network

Hang Zhao<sup>\*§</sup>, Yujing Wang<sup>\*§</sup>, Juanyong Duan<sup>\*</sup>, Congrui Huang<sup>\*</sup>, Defu Cao<sup>\*</sup>,  
Yunhai Tong<sup>†</sup>, Bixiong Xu<sup>\*</sup>, Jing Bai<sup>\*</sup>, Jie Tong<sup>\*</sup>, Qi Zhang<sup>\*</sup>

<sup>\*</sup>Microsoft, <sup>†</sup>Key Laboratory of Machine Perception, MOE, School of EECS, Peking University  
{hang.zhao, yujwang, juanyong.duan, conhua, t-decao, bix, jbai, jietong, qizhang}@microsoft.com, yhtong@pku.edu.cn

**Abstract**—Anomaly detection on multivariate time-series is of great importance in both data mining research and industrial applications. Recent approaches have achieved significant progress in this topic, but there is remaining limitations. One major limitation is that they do not capture the relationships between different time-series explicitly, resulting in inevitable false alarms. In this paper, we propose a novel self-supervised framework for multivariate time-series anomaly detection to address this issue. Our framework considers each univariate time-series as an individual feature and includes two graph attention layers in parallel to learn the complex dependencies of multivariate time-series in both temporal and feature dimensions. In addition, our approach jointly optimizes a forecasting-based model and a reconstruction-based model, obtaining better time-series representations through a combination of single-timestamp prediction and reconstruction of the entire time-series. We demonstrate the efficacy of our model through extensive experiments. The proposed method outperforms other state-of-the-art models on three real-world datasets. Further analysis shows that our method has good interpretability and is useful for anomaly diagnosis.

**Index Terms**—multivariate time-series, anomaly detection, graph attention network

### I. INTRODUCTION

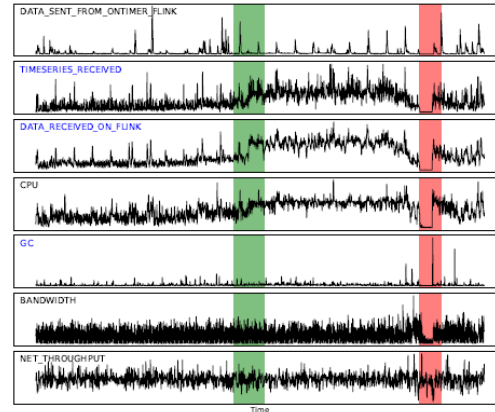


Fig. 1. An example of multivariate time-series input. Green indicates normal values and red indicates anomalies.

# Research Purpose

---

## ❖ Introduction

- 다변량 시계열 데이터는 하나의 센서 이상으로 인해 전체 시스템의 이상이 발생함 → 단변량 이상치 탐지가 아닌 다변량 이상치 탐지 방법론 필요
- 기존 시계열 이상치 탐지 연구들은 변수사이의 상관관계를 명시적으로 활용하지 않음
- 해당 연구에서는 Graph Attention Network(GAT)를 이용한 시계열 이상치 탐지 방법론 제안
- GAT를 활용해 feature 사이의 연관성, time사이의 연관성을 명시적으로 활용

# Time-series Anomaly Detection

---

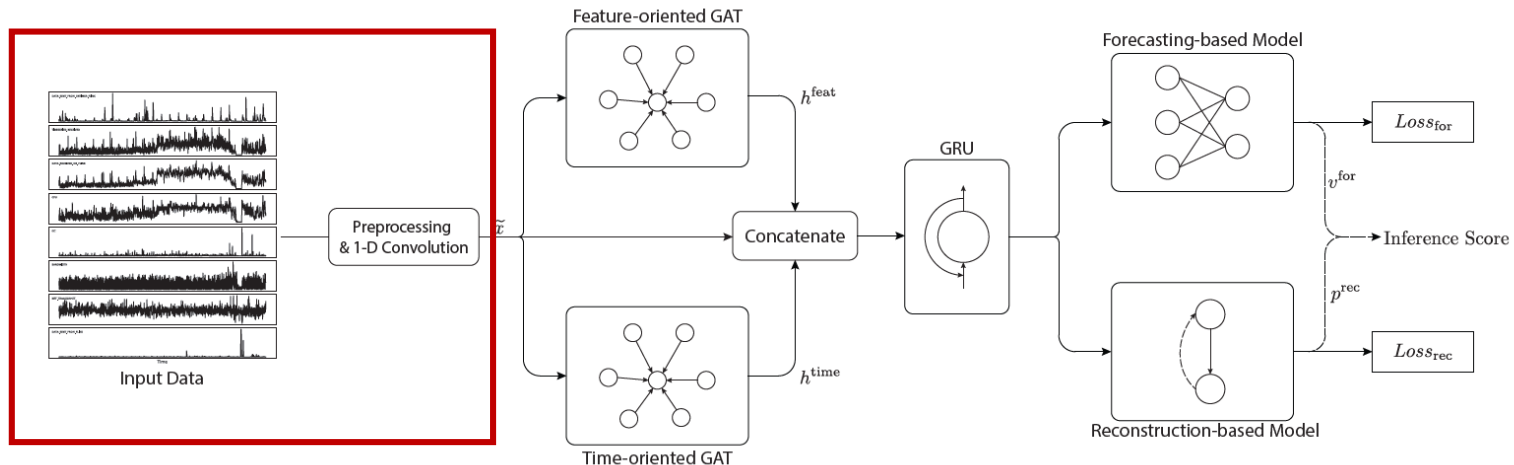
## ❖ Time-series Anomaly Detection

- **Forecasting-based:** 이전시점 데이터( $1 \sim t$ )를 통해 다음 시점 데이터( $t+1$ )를 예측, 정답값과 예측값이 일정량 이상 벌어지게 되면 이상이 발생한 것으로 판단
- **Reconstruction-based:** 입력된 전체 데이터를 모델을 통해 복원, 복원된 값을 기반으로 데이터의 이상 여부 판단
- 본 연구에서는 두가지 형태의 모델을 모두 활용하여 이상치 탐지에 활용

# MTAT-GAT

## ❖ Preprocessing

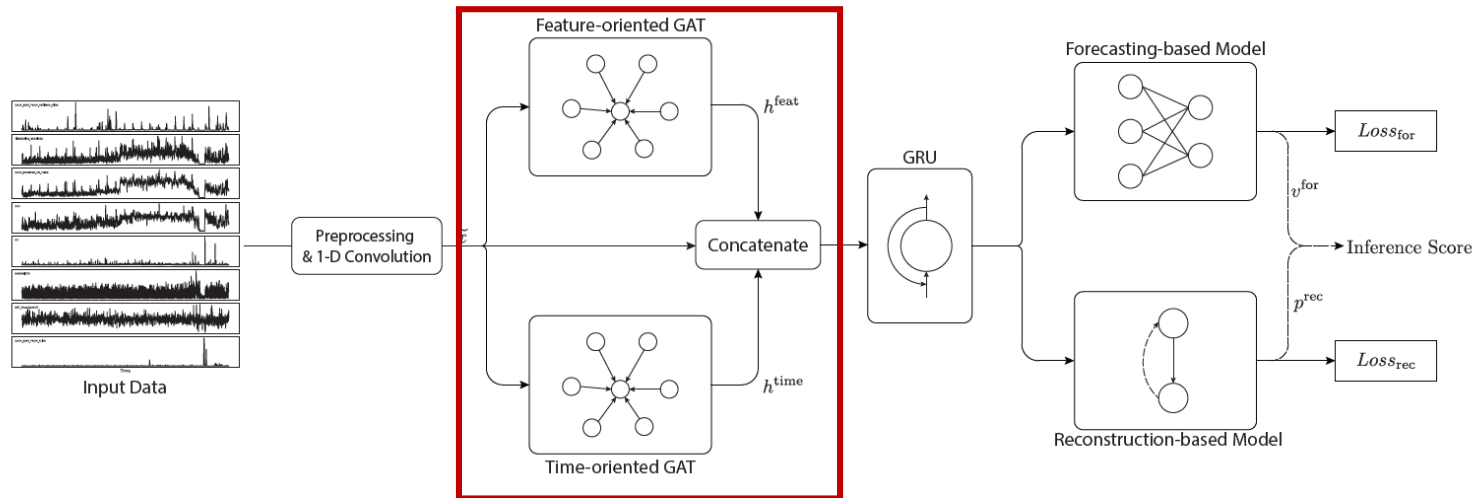
- Min-max scaling을 통해 데이터 정규화
- Training 데이터에 이상치가 존재할 경우 prediction-base, reconstruction-base 이상치 탐지 방법론은 취약
- Spectral Residual(SR)이라는 단변량 이상치 탐지 방법론을 활용해 이상치를 탐지한 후 해당 데이터를 정상 데이터 근방의 값으로 대체, SR은 가벼운 방법론으로 전처리에 활용 가능
- 최종적으로 1D-Conv를 활용해 기본적인 입력 데이터 구축



# MTAT-GAT

## ❖ Graph Attention

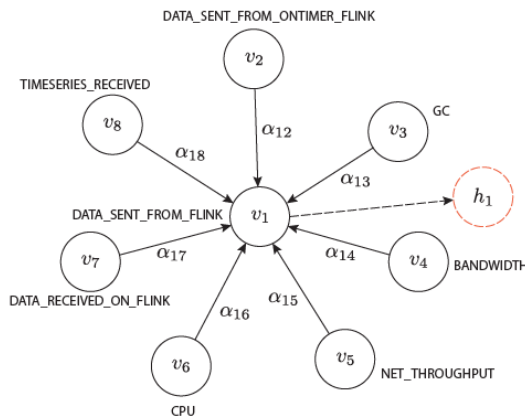
- 전처리된 데이터를 2개의 GAT layer에 입력
- Feature-oriented GAT: feature 사이의 상관관계를 학습하는 GAT
- Time-oriented GAT: time 사이의 상관관계를 학습하는 GAT



# MTAT-GAT

## ❖ Graph Attention

- 입력 그래프는 complete graph(모든 노드가 연결된 그래프)로 설정
- Feature-oriented GAT는 k(feature 개수)개의 노드가 있는 complete 그래프를 활용하며 각 노드의 dimension은 n(window size)  $\rightarrow$  output:  $k*n$
- Time-oriented GAT는 n개의 노드가 있는 complete 그래프를 활용, 각 노드의 dimension은 k  $\rightarrow$  output:  $n*k$
- GAT에서 사용한 attention mechanism 동일하게 활용



Feature-oriented GAT

$$e_{ij} = \text{LeakyReLU}(w^T \cdot (v_i \oplus v_j))$$
$$\alpha_{ij} = \frac{\exp(e_{ij})}{\sum_{l=1}^L \exp(e_{il})}$$

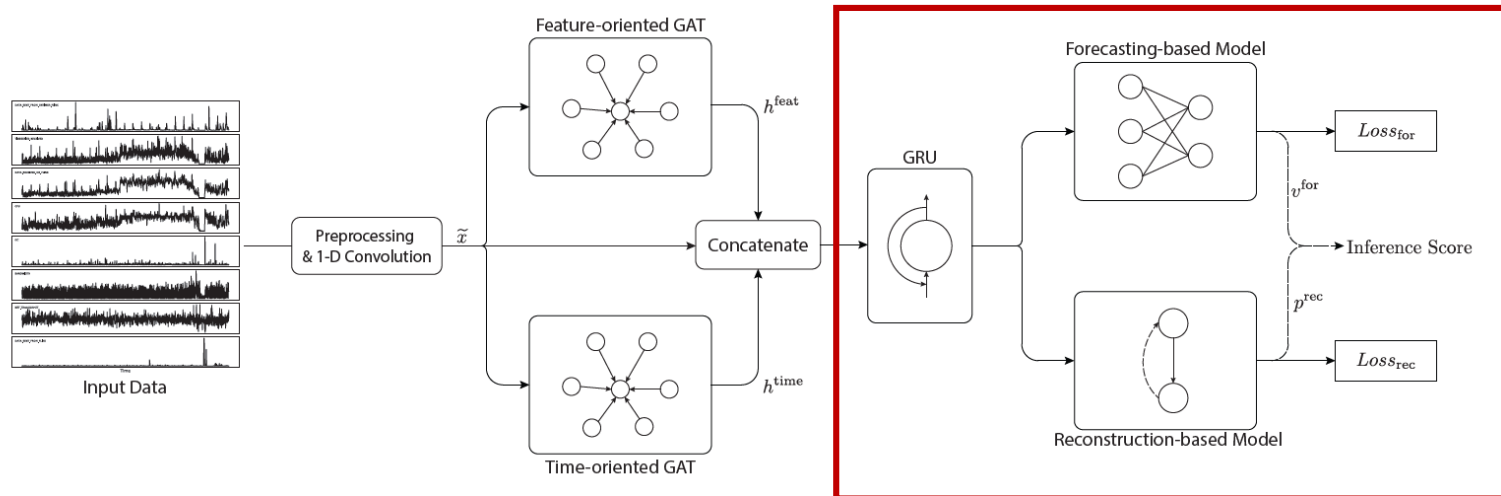
Attention Mechanism



# MTAT-GAT

## ❖ Anomaly Detection

- 기존 데이터, Feature-oriented GAT, Time-oriented GAT concatenate  $\rightarrow n*3k$
- GRU에 입력하여 feature 추출 후 Forecasting model과 Reconstruction model에 입력
- Forecasting model은 fully-connected neural network 활용 / RMSE를 loss로 활용
- Reconstruction model은 VAE(Variational Autoencoder) 활용 / VAE loss 활용
- 두가지 loss를 더해서 학습 진행



## ❖ Inference

- Forecasting model의 출력값( $\hat{x}$ ), Reconstruction model의 출력값( $p$ ) 활용해 anomaly score 계산
- Forecasting model의 예측값이 실제값과 차이가 커지면 anomaly score  $\uparrow$
- Reconstruction model의 출력값은 정상데이터일 확률로 생각할 수 있음,  $1-p$ : 이상데이터일 확률
- 감마는 하이퍼파라미터, 논문에서는 0.8을 활용

$$score = \sum_{i=1}^k s_i = \sum_{i=1}^k \frac{(\hat{x}_i - x_i)^2 + \gamma \times (1 - p_i)}{1 + \gamma}$$

# Experiments

## ❖ Comparison with SOTA

- 다양한 데이터셋에서 기존 방법론들에 비해 좋은 성능을 보임

TABLE III  
PERFORMANCE OF OUR MODELS AND BASELINES.

Method	SMAP			MSL			TSA		
	Precision	Recall	F1	Precision	Recall	F1	Precision	Recall	F1
Reconstruction based models									
OmniAnomaly	0.7416	0.9776	0.8434	0.8867	0.9117	0.8989	0.7028	0.8039	0.7499
KitNet	0.7725	0.8327	0.8014	0.6312	0.7936	0.7031	0.5579	0.8012	0.6577
GAN-Li	0.6710	0.8706	0.7579	0.7102	0.8706	0.7823	0.5302	0.7551	0.6229
MAD-GAN	0.8049	0.8214	0.8131	0.8517	0.8991	0.8747	0.5510	0.8284	0.6620
LSTM-VAE	0.8551	0.6366	0.7298	0.5257	0.9546	0.6780	0.6970	0.7736	0.7333
Forecasting based models									
LSTM-NDT	0.8965	0.8846	0.8905	0.5934	0.5374	0.5640	0.5833	0.7232	0.6457
DAGMM	0.5845	0.9058	0.7105	0.5412	0.9934	0.7007	0.5351	0.8845	0.6668
<b>MTAD-GAT</b>	0.8906	0.9123	<b>0.9013</b>	0.8754	0.9440	<b>0.9084</b>	0.6951	0.9352	<b>0.7975</b>

# Experiments

## ❖ Analyses

- Ablation study 결과 GAT 중에서는 feature-oriented GAT에 의한 성능 하락이 크게 발생
- Feature-oriented GAT attention score 시각화 결과 정상과 이상치 데이터에서 큰 차이를 보임

TABLE IV  
QUANTITATIVE RESULTS FOR ANALYSES. F1 SCORES ARE REPORTED.

Model	SMAP	MSL	TSA
<b>MTAD-GAT</b>	<b>0.9013</b>	<b>0.9084</b>	<b>0.7975</b>
w/o feature	0.8783	0.8851	0.7474
w/o time	0.8832	0.8897	0.7582
w/o prediction	0.8731	0.8857	0.7380
w/o reconstruction	0.8352	0.8058	0.7278

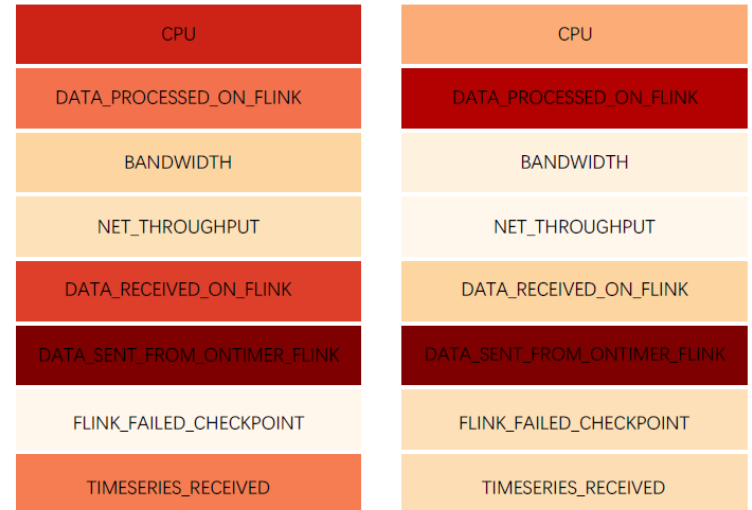


Fig. 5. Illustration of attention scores for feature *DATA\_SENT\_FROM\_FLINK* in the case of Figure 1. The left part visualizes the average attention scores at normal timestamps while the right part visualizes the attention scores when anomaly occurs. Darker color indicates higher attention scores.

*Thank You*