



Multivariate Time-series Anomaly Detection via Graph Attention Network(MTAD-GAT) (20- ICDM)

Hang Zhao, Yujing Wang, Juanyong Duan, Congrui Huang, Defu Cao, Yunhai Tong,
Bixiong Xu, Jing Bai, Jie Tong, Qi Zhang Microsoft,
Key Laboratory of Machine Perception, MOE, School of EECS, Peking University

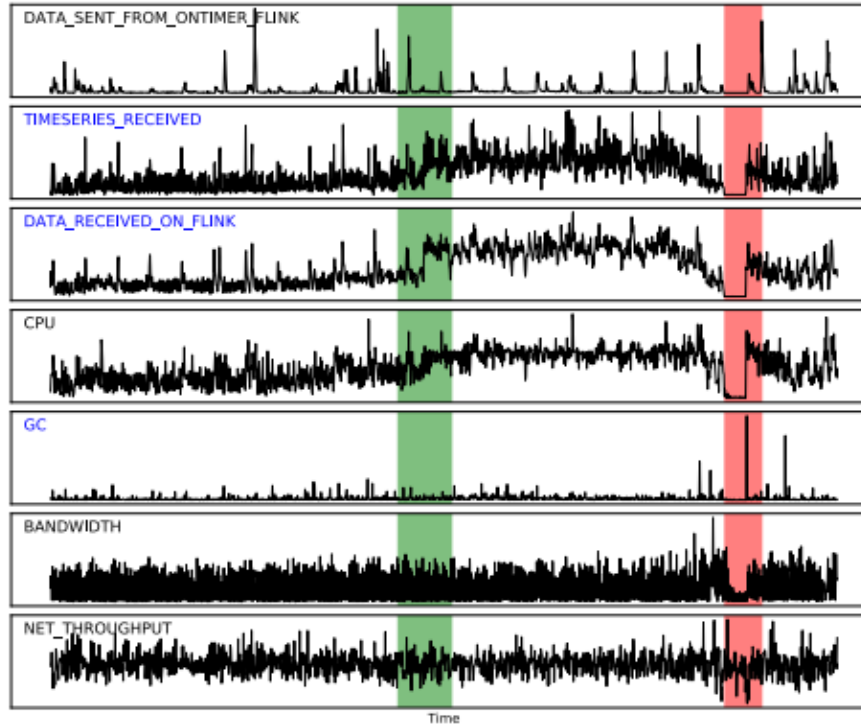
2024-08-13

Park Sae Joon
School of Computer Science and Engineering
Chung-Ang University
DMAIS Lab

Contents

- Previous Work
- MTAD-GAT
- Experiments
- Conclusions

■ Univariate time-series anomaly detection



■ Univariate time-series anomaly detection

■ Classic Methods

- 인위적으로 만든 features들을 사용해 정상 및 이상 패턴 모델링
- 통계적 기법이나 변환 기법을 활용하여 판단

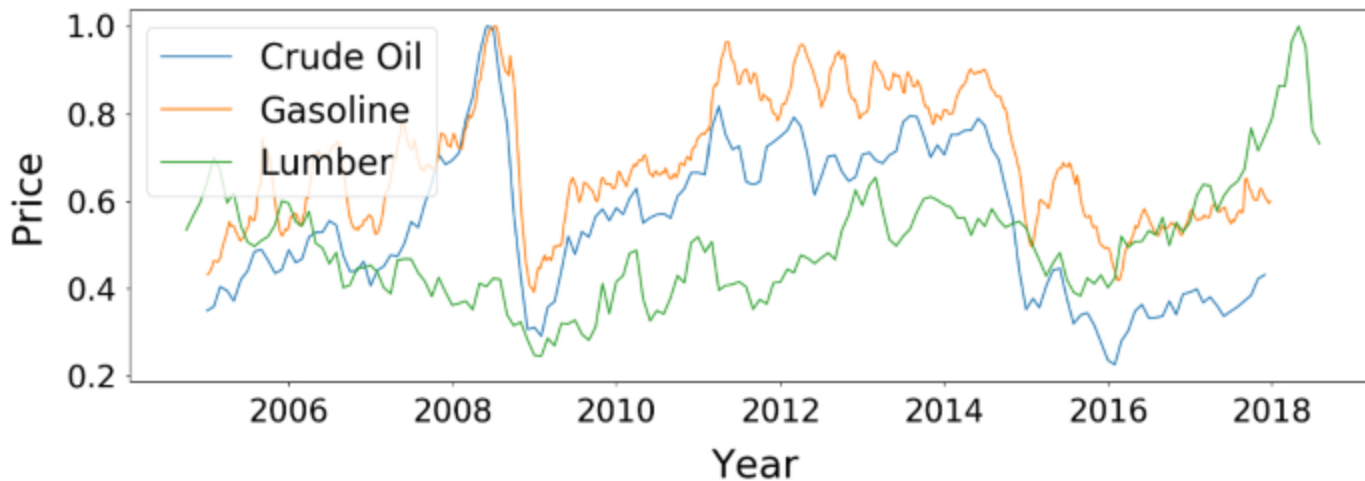
■ Recent Methods

- 신경망 기반의 방법을 활용해 이상 탐지에 사용
- 데이터의 복잡한 패턴을 학습할 수 있고, 고전적인 방법들 보다 뛰어난 성능 발휘 가능

■ Limit

- Multivariate time series에서 맞지 않음

- Multivariate time-series anomaly detection

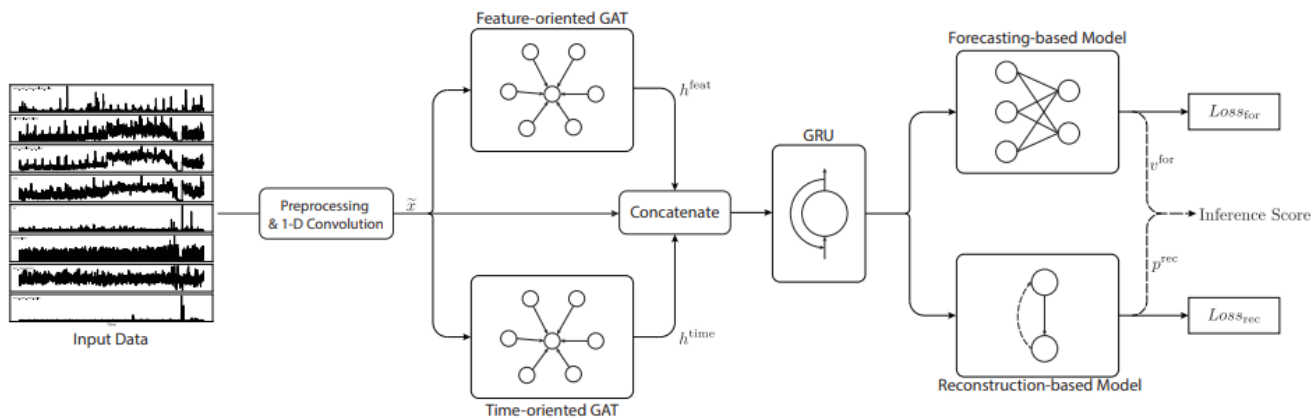


- **Multivariate time-series anomaly detection**
 - Forecasting-based Models
 - 미래의 시계열 값을 예측 및 실측값과의 오차를 기반으로 탐지
 - Reconstruction-based Models
 - 전체 시계열 데이터를 재구성해 실제 데이터 간의 차이를 기반으로 이상 탐지
 - 전체 시계열의 분포를 학습하고, 이를 재구성하는데 초점
 - Limit
 - 불규칙하거나 예측 불가능한 변동이 발생할 경우, 오탐 발생 가능(Forecast)
 - 단기적인 변동이나 특정 시점의 이상 탐지에 한계 존재 (Reconstruction)
 - 다변량 시계열 데이터의 특징 간 상관관계를 간과
 - 시간적 의존성 간과

■ Multivariate Time-series Anomaly Detection via Graph Attention Network

■ Solve previous works limits

- Graph Attention Network를 사용
- Feature oriented GAT를 이용해 특징 간 상호작용 명시적 학습
- Time oriented GAT를 이용해 시간적 의존성 학습



■ Data Preprocessing

■ Data

- 동일한 entity 형태의 단변량 시계열 데이터들이 다변량 시계열 데이터 형성

■ Data normalization

- 모든 시계열 데이터를 [0,1] 범위로 정규화

$$\tilde{x} = \frac{x - \min(X_{\text{train}})}{\max(X_{\text{train}}) - \min(X_{\text{train}})}$$

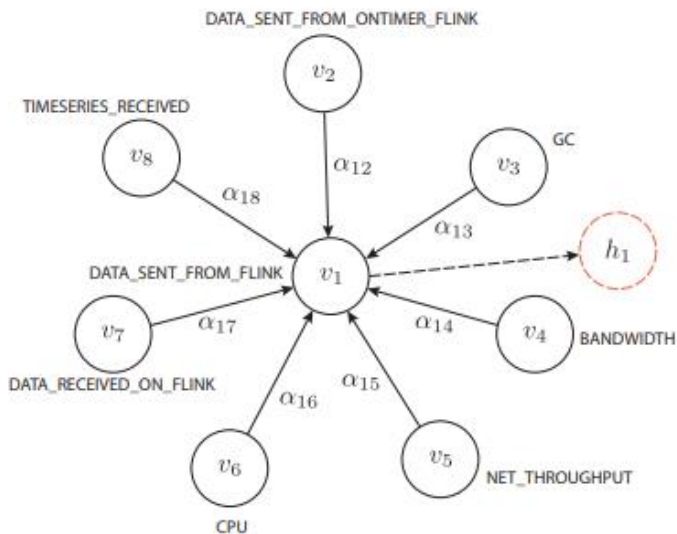
■ Data cleaning

- 예측 기반 모델과 재구성 기반 모델은 학습 데이터에서 irregular한 것에 민감
- SR방법으로 훈련 데이터의 이상점을 감지하고 제거
 - 가벼운 연산을 사용해 훈련 데이터 이상점 감지 후 주변의 정상 값으로 대체

■ Graph Attention

■ Feature oriented GAT Layer

- 다변량 시계열 데이터를 하나의 완전한 그래프로 간주



■ Graph Attention

■ Theorem

■ Output

$$h_i = \sigma\left(\sum_{j=1}^L \alpha_{ij} v_j\right)$$

■ Attention Score

$$e_{ij} = \text{LeakyReLU}(w^\top \cdot (v_i \oplus v_j))$$

$$\alpha_{ij} = \frac{\exp(e_{ij})}{\sum_{l=1}^L \exp(e_{il})}.$$

■ Graph Attention

■ Time oriented GAT Layer

- Timestamps들을 갖고 있는 sliding window를 complete한 graph로 간주
- Node x_t 는 timestamp t의 feature vector로 간주

■ GRU Layer

- GAT layer의 출력과 원본 시계열 데이터를 연결해 입력으로 사용
- 시계열 데이터의 연속적인 패턴들을 포착하는데 사용

- Joint Optimization
 - Forecasting-based Model + Reconstruction-based Model
 - 두 모델의 손실함수를 결합해 위 방법의 손실함수로 설정

$$Loss = Loss_{for} + Loss_{rec}$$

- **Forecasting Based Model**
 - 다음 timestamp의 값을 예측하는데 중점

$$Loss_{for} = \sqrt{\sum_{i=1}^k (x_{n,i} - \hat{x}_{n,i})^2}.$$

■ Reconstruction based Model

- 잠재 표현 z 에 대한 데이터의 주변 분포를 학습하는 것을 목표
- 손실함수의 앞부분은 input의 음의 log-likelihood
- 손실함수의 뒷부분은 정규화로 간주 가능

$$Loss_{rec} = -\mathbb{E}_{q_{\phi}(z|x)}[\log p_{\theta}(x|z)] + D_{KL}(q_{\phi}(z|x)||p_{\theta}(z))$$

■ Model Inference

■ Inference Score

- 모델은 예측 기반 오류와 재구성 기반 확률을 결합하여 이상여부 판단
- γ 는 비율을 조정하는 하이퍼 파라미터
- Score가 일정 임계값 초과하면 이상으로 간주
- 임계값은 Peak Over Threshold를 사용해서 자동으로 선택

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

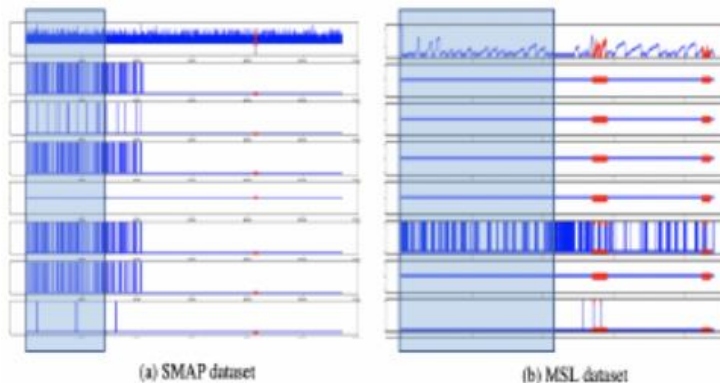
■ Datasets

- SMAP
- MSL
- TSA

	SMAP	MSL	TSA
Number of sequences	25	55	18
Training set size	135183	58317	39312
Testing set size	427617	73729	51408
Anomaly Rate(%)	13.13	10.27	10.58

■ Metrics

- Precision
- Recall
- F1-score
- AUC



■ Comparison with SOTAs

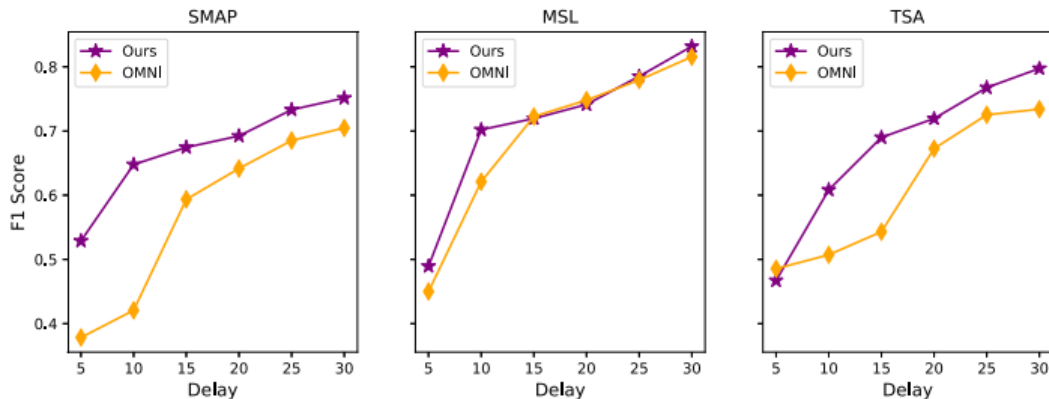
- 최신 모델들과의 비교
- 3가지 데이터 셋 모두에서 논문 모델이 가장 높은 F1-score를 기록

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
MTAD-GAT	0.8906	0.9123	0.9013	0.8754	0.9440	0.9084	0.6951	0.9352	0.7975

■ Evaluation with Different Delays

- 이상치가 지속적일 부분으로 생길 수 있기에 빠르게 detect하는게 중요
- 지연 메트릭에 따른 모델의 성능 추가로 평가
- $\delta=10$ 인 경우 기존 모델인 OmniAnomaly보다 약 54% 더 좋은 성능



■ Analyses

- Effectiveness of Graph Attention
 - 사용한 두개의 layer모두 성능에 중요한 기여
- Effectiveness of Joint Optimization
 - 사용한 두개의 모델을 결합한 것이 성능향상에 큰 기여

Model	SMAP	MSL	TSA
MTAD-GAT	0.9013	0.9084	0.7975
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

■ Analyses

■ Analysis of γ

- γ 값에 따른 성능 변화를 분석
- 0.8에서 가장 성능이 좋았지만 변화에 따라 성능 변동은 크지 않았음

γ	SMAP			MSL			TSA		
	Precision	Recall	F1	Precision	Recall	F1	Precision	Recall	F1
1.0	0.8832	0.9003	0.8917	0.8772	0.9413	0.9081	0.6874	0.9147	0.7849
0.8	0.8906	0.9123	0.9013	0.8754	0.9440	0.9084	0.6951	0.9352	0.7975
0.6	0.8814	0.9037	0.8924	0.8802	0.9315	0.9051	0.6861	0.9174	0.7851
0.4	0.8849	0.8991	0.8900	0.8817	0.9196	0.9003	0.6839	0.9233	0.7858

■ Analyses

■ Anomaly Diagnosis

■ Metrics

- HitRate

- NDCG

- Result

- 제안된 모델은 5개 내외의 상위 후보군 내에서 대부분 85% 이상의 정확도로 원인을 찾음

Model	NDCG@5	HitRate@100%	HitRate@150%
MTAD-GAT	0.8556	0.7428	0.8561

- GAT Layer를 사용해 다변형 시계열 데이터의 상관관계 명시적 학습
- Joint Optimization을 활용해 forecasting모델과 reconstruction모델 결합
- 기존의 모델들 보다 더 좋은 학습결과를 도출하고 이상 진단에도 좋은 성능을 보임
- 모델의 복잡성이 높고, 세가지 데이터셋만 사용했기에 추가적인 검증 연구가 필요

Thank You!

