

다변량 시계열 이상 탐지 과업에서 비지도 학습 모델의 성능 비교*

임 주 완,^{1*} 이 재 구^{2†}
^{1,2}국민대학교 (대학원생, 교수)

A Survey on Unsupervised Anomaly Detection for Multivariate Time Series*

Juwan Lim,^{1*} Jaekoo Lee^{2†}
^{1,2}Kookmin University (Graduate student, Professor)

요 약

다변량 시계열 이상 탐지 과업에서 정답 값이 존재하는 데이터를 얻는 것은 매우 시간 집약적인 일이다. 따라서 최근 정답 값이 필요 없는 비지도 학습법(unsupervised learning)에 관한 많은 연구가 진행되었다. 하지만 다변량 시계열 이상 탐지 과업에 특화된 주요 구조와 세부적인 특성에 대한 심화 있는 논의는 이루어지지 않았다. 본 논문에서는 비지도 학습 기반의 다변량 시계열 이상 탐지 모델과 특징점을 포괄적으로 분석하여 분류하였다. 전력 계통(power grid) 또는 Cyber Physical System(CPS)과 같은 현실 세계 데이터 집합에서 현실적인 이상 상황을 고려하여 학습을 진행하였고, 실험 결과를 바탕으로 각 모델의 정량적 성능을 비교 분석하였다. 성능 지표로는 정밀도(precision), 재현율(recall)과 F1 점수를 사용하여 성능을 측정하였다.

ABSTRACT

It is very time-intensive to obtain data with labels on anomaly detection tasks for multivariate time series. Therefore, several studies have been conducted on unsupervised learning that does not require any labels. However, a well-done integrative survey has not been conducted on in-depth discussion of learning architecture and property for multivariate time series anomaly detection. This study aims to explore the characteristic of well-known architectures in anomaly detection of multivariate time series. Additionally, architecture was categorized by using top-down and bottom-up approaches. In order to consider real-world anomaly detection situation, we trained models with dataset such as power grids or Cyber Physical Systems that contains realistic anomalies. From experimental results, we compared and analyzed the comprehensive performance of each architecture. Quantitative performance were measured using precision, recall, and F1 scores.

Keywords: Multivariate Time Series, Anomaly Detection, Unsupervised Learning, Deep Learning

Received(09. 20. 2022), Modified(12. 16. 2022),
Accepted(12. 29. 2022)

* 이 논문은 2022년도 정부(과학기술정보통신부)의 재원으로
정보통신기획평가원의 지원을 받아 수행된 연구임(No.RS-2022-00167194,미션 크리티컬 시스템을 위한 신뢰 가능한 인공지능)

* 이 논문은 2022년도 정부(과학기술정보통신부)의 재원으로
정보통신기획평가원의 지원을 받아 수행된 연구임(2022-0-01826, AI 기반 선도적 실전문제 해결 연구인재 양성)

† 주저자, dlacksdn2@kookmin.ac.kr

‡ 교신저자, jaekoo@kookmin.ac.kr(Corresponding author)

I. 서론

깊은 신경망 기반 심층 학습(deep learning)의 발전과 함께 이상 탐지 과업은 전력 계통(power grid)이나 Cyber Physical System(CPS) 등 비정상적 행위 감지가 필요한 여러 산업 분야에 많은 연구가 진행되었다[3,7,10,12,14]. 특히, 시계열 데이터 기반 이상 탐지 과업은 정상 시계열 패턴(pattern)에서 벗어난 비정상 데이터 또는 패턴을 감지하는 것을 의미한다.

시계열 데이터는 대표적으로 두 가지로 분류할 수 있다. 첫 번째는 단변량(univariate) 시계열로 Fig. 1. (a)와 같이 시간 순서에 따라 나열된 1차원 관측값들이다. 두 번째는 다변량(multivariate) 시계열로, Fig. 2. (b)에서 볼 수 있듯 복수의 단변량 시계열 관측값들을 같은 시간 단위로 묶은 형태이다.

심층 학습 이전의 전통적인 시계열 이상 탐지 방법으로 선형 분류기, Support Vector Machine(SVM)[1]를 이용한 정상/비정상 분류 방법과 K-Nearest Neighbor(KNN)처럼 패턴의 군집 간 유사도 측정 방법이 있다. 그러나 전통적인 방법은 다변량 시계열으로 변화하고 비정상 패턴이 복잡해짐에 따라 차원의 저주(curse of dimensionality) 문제에 직면하는 한계점을 갖는다[2].

지난 10여 년 동안 심층 학습은 자연어 처리나 사진이나 영상 분석, 시계열 데이터 예측 등 대부분의 인공지능 과업에서 비약적인 성장 향상을 이루었다. 깊은 신경망 모델의 특징 추상화를 통해 복잡하고 많은 양의 데이터에 대한 표현 학습이 가능하기 때문이다. 하지만, 일반적으로 정답 값이 필요한 지도 학습(supervised learning)은 시계열 이상 탐지 과업에 매우 비효율적이다. 시계열 데이터의 특성상 메시지 순서에 따른 관측값에 정상 또는 이상 여부를 판

단하여 정답 값을 부여하는 것은 매우 비효율적인 노동집약적 일이며 해당 분야에 대한 지식을 갖춘 전문가가 필요하기 때문이다. 따라서 최근 정답 값이 필요 없는 비지도 학습법(unsupervised learning)에 대한 많은 연구가 진행되었다[3,5,7,10,12,14,17].

본 연구에서는 다변량 시계열 이상 탐지 과업에서 최신 비지도 학습 기반 모델을 살펴보고, 모델의 구조에 따른 특징점을 분석한다. 나아가 현실 공공 데이터 집합에 대한 정량적 성능 비교와 분석을 제시한다. 특히, 모델의 구조를 하향식(top-down)과 상향식(bottom-up) 양 접근을 통해 세부적 특성까지 세분화하였다. 또한, 현실 세계의 실질적 이상 데이터 또는 잡음(noise)의 존재를 고려하여 모델의 포괄적인 성능을 측정하였다. 정량적 지표로는 정밀도(precision)와 재현율(recall), F1 점수를 이용하여 이상 데이터 탐지 성능을 측정하였다.

II. 모델 설명

Fig 2.에서 제시된 것처럼 모델 세분화는 하향식 접근으로 주 역할을 하는 구조(main architecture) 분류와 상향식 접근으로 보조적인 특성을 구분하였다. 또한, Zhigan Li et al.[10] 연구에서 학습 집합에 이상 데이터가 존재함을 고려하여 데이터를 정제하는 '사전 필터링(pre-filtering)'을 사용하는 모델도 구분하였다.

2.1 AutoEncoder(AE)

대표적인 이상 탐지 모델인 AutoEncoder(AE)는 인코더(encoder)를 통해 입력값을 저차원의 잠재 변수로 압축하고 이를 다시 디코더(decoder)로 복원한다. 학습 시에 정상 데이터만을 사용하면 AE는 임의의 모든 입력값에 대해 정상 데이터로 복원하는 경향을 보인다. 따라서 추론 시에 입력값과 복원값의 유사도 차이를 통해 데이터의 이상 여부를 판단한다. AE 구조를 사용하면 정답 값이 없어도 정상 데이터만으로 학습이 가능한 장점이 있다.

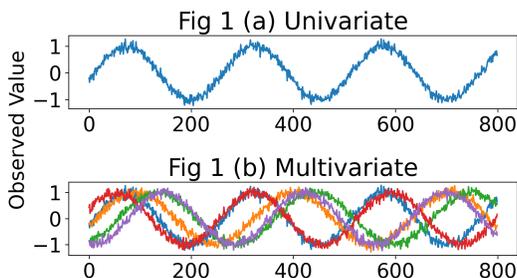


Fig. 1. Examples of time series (a) Univariate, (b) Multivariate

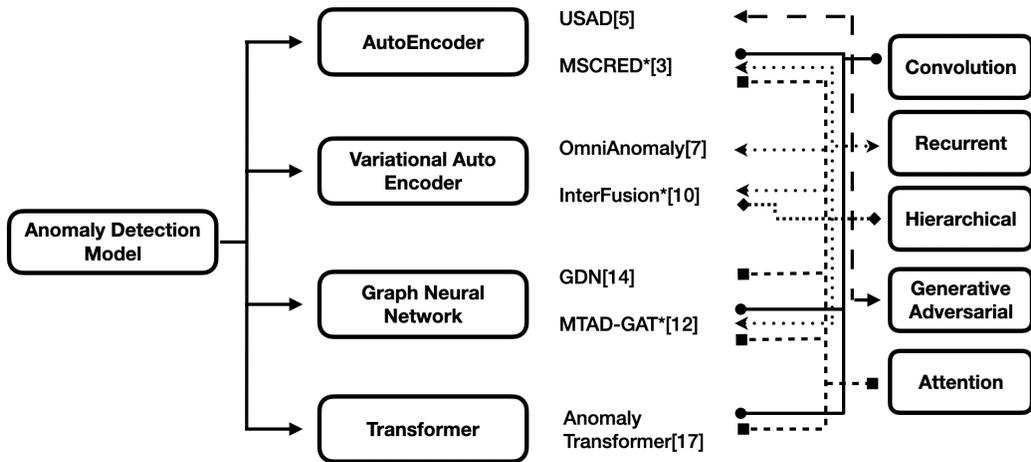


Fig. 2. Taxonomy of Multivariate Time Series Unsupervised Anomaly Detection. * denotes models using pre-filtering strategy

2.1.1 MSCRED[3]1)

MSCRED[3]는 Fig 3.에서 보인 것처럼 AE 구조에 합성곱 신경망과 Convolutional Long Short Term Memory(Conv-LSTM)(4)을 사용한다. 일정한 시간 간격(window) 내에서 'signature matrix'를 생성하고 이를 합성곱 신경망 인코더 및 디코더에 통과시켜 단변량 시계열 간의 상관관계를 학습한다. 동시에 각 합성곱 신경망의 출력값을 Conv-LSTM에 통과시켜 시간 정보를 학습한다. signature matrix는 다수의 정상 데이터들로 인해 데이터 내에 존재하는 잡음의 영향력이 줄어 사전 필터링의 효과를 가진다.

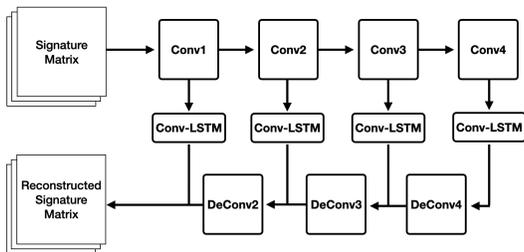


Fig. 3. MSCRED(3) architecture

2.1.2 USAID[5]2)

USAID[5]는 AE 구조에 적대적 학습 기법 (adversarial training)(6)을 적용한 모델로, 하나의 인코더를 공유하는 두 개의 디코더로 이루어졌다. Fig. 4.에서 볼 수 있듯이 인코더와 디코더1로 이루어진 AE1은 입력값을 압축한 뒤 다시 원래의 입력값과 유사하게 복원하도록 학습한다. 한편, 인코더와 디코더2로 이루어진 AE2는 AE1이 복원한 값과 원래의 입력값을 구별하도록 학습한다. 이는 AE2에서의 디코더가 추론 시에 정상 데이터와 매우 유사한 이상 데이터까지도 구별하도록 학습하기 위함이다.

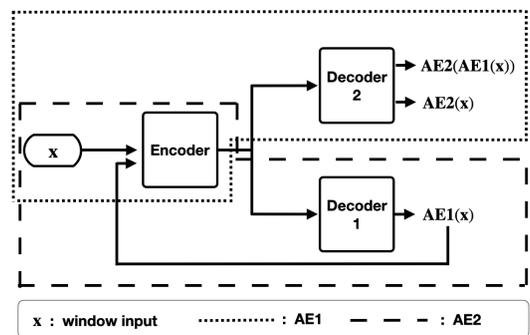


Fig. 4. USAID(5) architecture

1) <https://github.com/7fantasysz/MSCRED>

2) <https://github.com/manigalati/usad>

2.2 Variational AutoEncoder(VAE)

Variational AutoEncoder(VAE)는 기본 (vanilla) AE와 다르게, 데이터가 생성되는 분포를 학습한다. AE는 입력값을 저차원의 고정 크기 벡터로 압축하는 반면, VAE는 입력값을 저차원의 분포로 압축하고, 이 분포로부터 새로운 데이터를 생성할 수 있다. 따라서 입력값이 정상 데이터로 생성되었을 확률을 통해 이상 데이터를 감지한다. 시계열 데이터와 같이 비결정적인 데이터에 대한 효과적인 모형화가 가능하다는 장점이 있다.

2.2.1 OmniAnomaly[7]³⁾

Fig. 5.의 OmniAnomaly[7]은 시간 정보를 학습하기 위해 인코더와 디코더를 순환 신경망인 Gate Recurrent Unit(GRU)[8]로 구성된 모델이다. GRU를 통해 시계열 입력값을 평균과 편차를 갖는 저차원의 분포로 압축하고, 이 분포로부터 새로운 데이터를 만들어낸다. 기존 VAE는 분포를 가우스 분포와 같은 간단한 분포로 가정한다. 하지만 시계열 데이터를 생성하는 분포는 가우스 분포보다 더 복잡한 형태를 보인다. 따라서, OmniAnomaly[7]은 복잡한 분포를 모형화하기 위해 planar normalizing flow[9]기법을 적용했다.

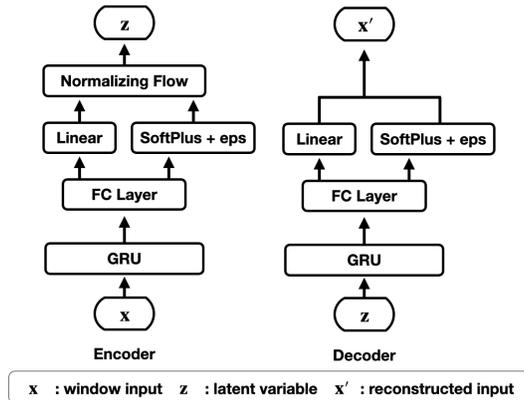


Fig. 5. OmniAnomaly(7) architecture

2.2.2 InterFusion[10]⁴⁾

InterFusion[10]은 단변량 시계열 간의 상관관계와 시간 정보를 동시에 학습하기 위해 계층 구조로 이루어진 Hierarchical VAE(H-VAE)[10]를 이용한다. H-VAE는 2가지 관점에서의 임베딩(embedding)을 진행하는데, 첫 번째는 시간 정보 임베딩이고, 두 번째는 단변량 시계열 간의 상관관계 임베딩이다. 시간 정보 임베딩은 1차원 합성곱 인코더 및 디코더를 통해 사전 필터링을 적용함과 동시에 시간 정보를 학습하는 것을 의미한다. 한편, 상관관계 임베딩은 1차원 디코더를 통해 복원된 값을 GRU와 normalizing flow 기반의 Real-Non Volume Preserving(NVP)[11]를 사용하여 저차원의 분포로 압축하는 것을 의미한다. 전체적인 구조는 OmniAnomaly[7]와 유사하지만 사전 필터링의 여부와 계층 구조의 적용 차이가 있다.

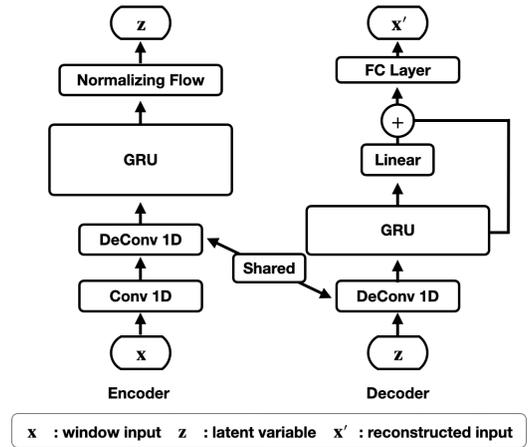


Fig. 6. InterFusion[10] architecture

2.3 Graph Neural Network(GNN)

시계열 데이터에서 각 단변량 시계열들의 상관관계를 학습하기 위해, 데이터의 구조를 그래프로 변환할 수 있다. 이를 Graph Neural Network(GNN)에 적용하면, 단변량 시계열 간의 상관관계와 시간 정보를 명시적으로 동시에 학습할 수 있는 장점이 있다.

3) <https://github.com/NetManAIOps/OmniAnomaly>

4) <https://github.com/zhlee/InterFusion>

2.3.1 MTAD-GAT[12]⁵⁾

MTAD-GAT[12]는 1차원 합성곱과 Spectral Residual(SR)[13]을 사용한 데이터 정제 과정을 통해 사전 필터링을 진행한다. 이후 각 단변량 시계열 간의 상관관계를 학습하는 특징 기반 집중(feature oriented attention), 각 단변량 시계열의 시간 정보를 학습하는 시간 정보 기반 집중(time oriented attention), 그리고 1차원 합성곱의 출력값을 합친 뒤 GRU의 입력값으로 사용한다.

또한 이상치를 판단하는 방법으로 예측 모델과 복원 모델을 동시에 활용하는데, 예측 모델은 완전 연결 층으로 구성했고 다음 시점의 예측값과 실제값의 차이를 통해 이상 데이터를 감지한다. 반면 복원 모델은 VAE로 구성했고 입력값이 잘 복원되었을 확률을 통해 이상 데이터를 감지한다.

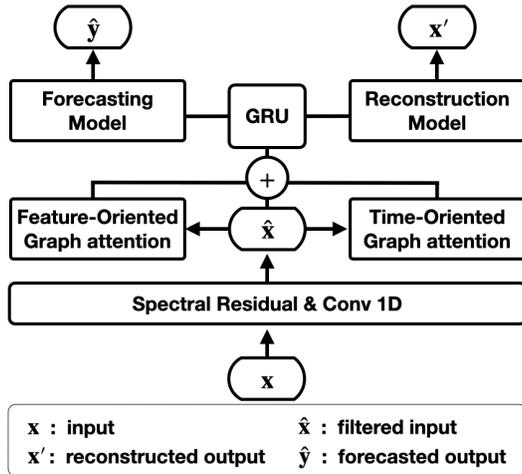


Fig. 7. MTAD-GAT[12] architecture

2.3.2 GDN[14]⁶⁾

GDN[14]은 특징 임베딩(sensor embedding) 단계에서 다변량 시계열의 각 차원을 하나의 노드(node) 벡터로 모형화한다. 이후, 그래프 구조 학습(graph structure learning) 단계에서 노드 간의 유사도[15]를 계산하여 인접행렬을 생성하고 그래프 어텐션 기반 예측(graph attention-based forecasting) 단계에서 인접행렬과 노드 벡터를

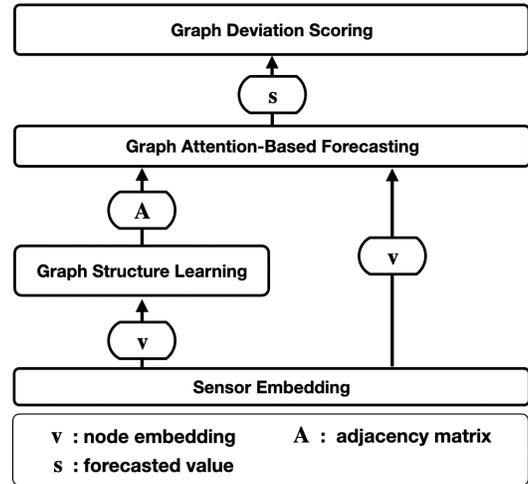


Fig. 8. GDN[14] architecture

사용하여, 각 단변량 시계열의 다음 시점을 예측한다. 이때 각 단변량 시계열마다 예측값과 실제값의 차이를 계산하고, 가장 큰 차이 값을 이상 점수로 사용한다. 이상 점수가 임의의 임계 값을 초과하는 경우, 이상 데이터로 판단한다.

2.4 Transformer(TF)

Transformer[16]는 attention을 통해 기존 순환 신경망 기반 모델들이 갖는 정보 소실 문제를 극복한 방법론이다. 순환 신경망은 이전 시점의 은닉 상태를 계산해야 다음 시점의 은닉 상태를 계산할 수 있어 병렬 연산할 수 없는 단점이 있다. 반면, Transformer는 모든 입력값을 토큰화(tokenize)한 후 동시에 연산하므로 병렬적으로 계산이 가능한 장점이 있다.

2.4.1 Anomaly Transformer[17]⁷⁾

Anomaly Transformer[17]는 기본적인 Transformer 구조에 prior association이라는 새로운 네트워크를 정의하여 추가한 모델이다.

prior association은 지역적인 연관성을 의미하며 이상 데이터가 인접 시점 데이터들과 높은 연관성을 갖는다는 사전 지식을 이용한 것이다.

한편 series association은 이상 데이터가 드물

5) <https://github.com/ML4ITS/mtad-gat-pytorch>

6) <https://github.com/d-ailin/GDN>

7) <https://github.com/thuml/Anomaly-Transformer>

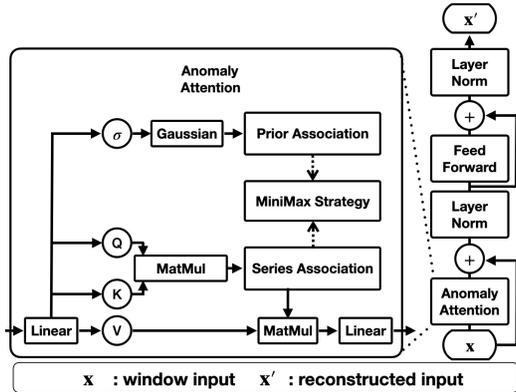


Fig. 9. Anomaly Transformer(17) architecture

게 존재한 점을 이용하여 전체적인 시점에 존재하는 데이터들과의 연관성이 낮다는 점을 이용한다. prior association을 표현하기 위해 가우스 분포에 학습 가능한 매개변수 σ 를 사용해서 인접 시점들과의 연관성을 계산한다. 한편, 한 시점을 기준으로 기본 transformer의 집중(attention) 연산과 동일한 연산을 통해 series association을 구한다. 또한 정상 데이터와 이상 데이터의 차이를 극대화하기 위하여 prior association과 series association의 차이를 극대화하는 MiniMax 전략(17)을 사용하는 것이 특징이다.

III. 성능 평가

3.1 데이터 집합

모든 데이터 집합은 학습 집합에 정상 데이터만을

Table 2. Dataset description

Dataset	Train	Test	#Dim.
SWaT	495,000	449,919	51
WADI	1,209,601	172,801	123
SMAP	135,183	427,617	25

포함하고 시험 집합만 이상 데이터가 존재한다.

Secure Water Treatment(SWaT)(18)은 하수 처리장의 이상 상태 감지 목적을 위해 고안된 시험 환경(testbed)에서 측정된 51개의 센서 차원을 갖는 다변량 시계열 데이터 집합이다. 시험 집합에 대략 12.14%의 이상 데이터가 존재한다.

Water Distribution(WADI)(19)는 SWaT 데이터 집합에서 확장된 개념의 다변량 시계열 데이터 집합으로, 총 123개의 센서 차원을 갖는다. 시험 집합에는 대략 5.7%의 이상 데이터가 존재한다.

Soil Moisture Active Passive(SMAP)(20)는 NASA에서 수집한 인공위성의 토양 수분 관측 데이터로, 총 1개의 신호와 24개의 원-핫 인코딩(one-hot encoded) 처리된 센서 차원으로 이루어진 데이터 집합이다. 시험 집합에는 대략 12.79%의 이상 데이터가 존재한다.

Mars Science Laboratory(MSL)(21)은 NASA의 화성 탐사 로봇 데이터 집합으로 총 1개의 신호와 54개의 원-핫 인코딩 처리된 센서 데이터 집합이다. 시험 집합에는 대략 10.53%의 이상 데이터가 존재한다.

본 논문에서는 학습 집합에도 잠재적인 이상 데이터 또는 잡음이 존재한다는 현실적인 조건의 실험을 진행하기 위해 다음과 같은 두 가지 추가 실험을 진

Table 1. Experiment results on four public original dataset. Underline is the best performance in main architecture. Bold is the overall best performance. * denotes models using pre-filtering method. OOM denotes out of memory

Main arch.	DATASET Model	SWaT			WADI			SMAP			MSL		
		P	R	F1									
AE	MSCRED*(3)	0.944	0.886	<u>0.914</u>	OOM			0.928	0.524	0.670	0.774	0.752	0.748
	USAD(5)	0.947	0.85	0.896	0.600	0.590	0.595	0.964	0.558	<u>0.707</u>	0.765	0.908	<u>0.830</u>
VAE	InterFusion*(10)	0.947	0.909	<u>0.928</u>	0.819	0.897	<u>0.856</u>	0.698	0.851	<u>0.767</u>	0.895	0.927	<u>0.911</u>
	OmniAnomaly(7)	0.976	0.690	0.808	0.774	0.654	0.709	0.946	0.554	0.699	0.915	0.869	0.891
GNN	MTAD-GAT*(12)	0.926	0.892	<u>0.909</u>	0.767	0.329	0.460	0.846	0.999	<u>0.916</u>	0.962	0.967	0.964
	GDN(14)	0.997	0.647	0.785	0.810	0.350	<u>0.489</u>	0.132	0.931	0.231	0.201	0.640	0.306
TF	Anomaly Transformer(17)	0.948	0.968	0.958	0.910	1.000	0.953	0.938	0.995	0.965	0.922	0.987	0.953

Table 3. Experiment results on four public dataset with random perturbation. Underline is the best performance in main architecture. Bold is the overall best performance. * denotes models using pre-filtering method. OOM denotes out of memory

Main arch.	DATASET Model	SWaT			WADI			SMAP			MSL		
		P	R	F1									
AE	MSCRED*[3]	0.732	0.976	<u>0.837</u>	OOM			0.466	0.906	0.615	0.747	0.744	0.746
	USAD[5]	0.945	0.671	0.785	0.096	1.000	0.176	0.528	0.994	<u>0.689</u>	0.953	0.652	<u>0.774</u>
VAE	InterFusion*[10]	0.886	0.936	<u>0.910</u>	0.978	0.320	<u>0.483</u>	0.828	0.657	<u>0.733</u>	0.936	0.886	<u>0.910</u>
	OmniAnomaly[7]	0.121	0.999	0.217	0.057	0.999	0.108	0.962	0.546	0.697	0.916	0.854	0.884
GNN	MTAD-GAT*[12]	0.988	0.779	<u>0.871</u>	0.998	0.278	0.435	0.758	0.859	<u>0.805</u>	0.903	0.919	0.911
	GDN[14]	0.932	0.623	0.747	0.813	0.305	<u>0.444</u>	0.129	0.897	0.225	0.195	0.454	0.273
TF	Anomaly Transformer[17]	0.915	1.000	0.956	0.775	1.000	0.873	0.936	0.860	0.896	0.905	0.859	0.881

행하였다. 첫 번째로 잡음에 대한 각 모델의 강건성을 확인하기 위해 학습 집합에 일정 시간 간격으로 잡음을 추가하여 실험하였다. 두 번째로, 실제 이상 데이터를 포함하여 학습했을 때, 각 모델의 성능을 비교하기 위해 원본 데이터의 시험 집합을 6:2:2 비율로 나누어 다시 학습, 검증, 시험 집합으로 분배하여 실험하였다. 분배 시에는 시간 정보의 손실을 막기 위해 무작위로 추출하거나 데이터를 섞지 않았으며 이상 데이터의 비율보다는 시계열 데이터 길이 비율에 중점을 두어 분배하였다.

3.2 성능 지표

Table 4.은 정밀도와 재현율을 구하기 위한 혼동 행렬이다. 이상 탐지에서 주로 사용하는 성능 측정 지표는 F1 점수이다. F1 점수는 정밀도 (precision)와 재현율(recall)의 조화 평균으로서, 정밀도와 재현율이 trade-off 관계이기 때문에 이들을 보완한 종합적인 성능 지표로 사용할 수 있다. 또한 F1 점수는 데이터의 클래스 불균형 문제가 있더라도 이와 상관없이 적용 가능한 평가 지표이다. F1 점수의 계산 수식은 다음과 같다.

$$\text{Precision} = \frac{TP}{TP + FP} \quad (1)$$

Table 4. Confusion matrix

		Label	
		True	False
Prediction	Positive	TP	FP
	Negative	FN	TN

$$\text{Recall} = \frac{TP}{TP + FN} \quad (2)$$

$$\text{F1 Score} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \quad (3)$$

3.3 실험 결과

Table 1.과 Table 2.에서 볼 수 있듯, 데이터 집합에 잡음을 추가하였을 때 MSCRED[3]의 F1 점수는 SWaT 데이터 집합에서 최대 0.077 하락하였고 USAD는 0.111 하락하였다. 이는 signature matrix를 활용하여 각 단변량 시계열 사이의 상관 관계를 고려함과 동시에 잡음의 영향력을 완화했기 때문으로 볼 수 있다. 단, SMAP와 MSL 데이터 집합은 하나의 특징을 제외한 나머지 특징들이 원-핫 인코딩된 특징들이기 때문에 MSCRED[3]의 signature matrix의 역할이 줄어들어, 오히려 단순한 완전 연결층 구조를 갖는 USAD[5]가 더 높은 F1 점수를 달성한 것으로 해석할 수 있다.

WADI 데이터 집합에서는 전체 데이터의 양과 특징 차원이 매우 커 MSCRED[3]에서 signature matrix를 생성하는 것이 불가능하였다.

VAE 모델 중에선 전반적으로 InterFusion[10]의 F1 점수가 OmniAnomaly[7]보다 높았다. 특히, 학습 집합에 잡음이 추가된 상황에서 WADI 데이터 집합에 대한 InterFusion[10]의 F1 점수가 0.374 하락하였지만, OmniAnomaly[7]은 F1 점수가 0.601 하락하며, InterFusion[10]에 비해 잡음에 대한 대처 능력이 부족한 것으로 해석할 수

Table 5. Experiment results on four public dataset with real anomalies. Underline is the best performance in main architecture. Bold is the overall best performance. * denotes models using pre-filtering method.

Main arch.	DATASET Model	SWaT			WADI			SMAP			MSL		
		P	R	F1									
AE	MSCRED*[3]	0.804	0.608	<u>0.693</u>	1.000	0.72	<u>0.837</u>	0.584	0.411	<u>0.482</u>	0.774	0.741	<u>0.757</u>
	USAD[5]	0.331	0.915	0.145	0.556	0.231	0.327	0.159	0.958	0.272	0.132	0.916	0.231
VAE	InterFusion*[10]	0.729	1.000	<u>0.843</u>	0.931	0.803	<u>0.862</u>	0.52	0.746	0.613	0.691	0.984	0.812
	OmniAnomaly[7]	0.79	0.697	0.741	0.101	0.999	0.184	0.511	0.88	<u>0.647</u>	0.122	0.999	0.217
GNN	MTAD-GAT*[12]	0.727	0.918	<u>0.812</u>	0.999	0.251	0.401	0.777	1.000	<u>0.875</u>	0.623	0.299	<u>0.404</u>
	GDN[14]	0.970	0.633	0.766	0.600	0.439	<u>0.507</u>	0.128	0.760	0.219	0.095	0.321	0.152
TF	Anomaly Transformer[17]	0.845	1.000	0.916	0.899	1.000	0.947	0.918	0.996	0.955	0.791	0.755	0.773

았다.

GNN 모델에선 MTAD-GAT[12]가 GDN[14] 보다 전반적으로 더 높은 F1 점수를 달성했다. 특히 SMAP와 MSL 데이터 집합에서 GDN[14]의 F1 점수는 MTAD-GAT[12]에 비해 매우 낮은 점수를 갖는다. 이는 GDN[14]의 특성상 원-핫 인코딩 형태의 시계열 간 노드 유사도는 증가하고 나머지 1개의 신호 시계열과의 노드 유사도는 감소하여, 중요한 신호 차원에 대한 학습이 제대로 이루어지지 않기 때문이라고 해석할 수 있다. 한편, 잡음이 추가된 상황에서, MTAD-GAT[12]와 GDN[14] 모두 F1 점수 하락 폭이 다른 구조의 모델들보다 비교적 적었다.

Anomaly Transformer[17]의 경우, Table 1.과 Table 2.에서 볼 수 있듯, MSL 데이터 집합을 제외한 나머지 데이터 집합에서 가장 높은 성능을 보였다. 또한 잡음이 추가된 상황에서 다른 모델들과 비교했을 때 F1 점수의 하락 폭은 유사했지만, 잡음을 추가하지 않은 데이터로 학습한 다른 모델들보다 높은 성능을 보였다. 특히 Table 1.과 Table 2.에서 SWaT과 WADI 데이터 집합에 대한 Recall으로 대부분 1.00 또는 이에 근사한 값을 얻었다. 이는 Anomaly Transformer[17]가 데이터에 대한 이상치를 판단할 때, 입력값과 복원값의 차이보다는 prior association과 series association의 차이에 매우 의존적이기 때문이다. 즉, 정상 데이터인 시점일 때도 association 간의 차이가 이상 데이터 시점과 유사하다면, 입력값과 복원값의 차이와 관계없이 이상 데이터로 판단하는 경향이 있다.

Table 5.는 각 데이터의 시험 집합을 다시 학습,

검증, 시험 집합으로 나누어 학습 집합에 실제 이상 데이터가 존재했을 때의 실험한 결과이다. Table 6.에서 볼 수 있듯, USAD[5]의 전체 데이터 집합에 대한 평균 F1 점수는 0.244로 다른 모델들에 비해 제일 낮은 평균 성능을 기록하였다. 이는 학습 단계에서 USAD[5]가 실제 이상 데이터를 학습하였을 때 매우 취약한 것으로 해석할 수 있다.

반면, MSCRED[3]는 다변량 시계열의 상관관계를 나타내는 signature matrix를 생성하면서, 사전 필터링을 수행한다. 평균 F1 점수는 같은 AE 모델인 USAD[5]보다 0.448 높은 0.692이다. 이는 MSCRED[3]가 사전 필터링 과정을 거치면서 시간 정보까지 학습할 수 있는 Conv-LSTM을 사용하였기 때문에 USAD[5]보다 더 높은 성능을 보인 것으로 해석할 수 있다.

OmniAnomaly[7]와 InterFusion[10]의 가장 큰 차이점은 사전 필터링의 유무이다.

Table 6. Average F1 score and standard deviation on four datasets. * denotes models using pre-filtering method.

Main arch.	Model	avg. F1	std.
AE	MSCRED*[3]	<u>0.692</u>	0.0173
	USAD[5]	0.244	0.0044
VAE	InterFusion*[10]	<u>0.783</u>	0.0098
	OmniAnomaly[7]	0.447	0.0621
GNN	MTAD-GAT*[12]	<u>0.623</u>	0.0491
	GDN[14]	0.411	0.0597
TF	Anomaly Transformer[17]	0.898	0.0054

OmniAnomaly[7]는 시계열 데이터를 바로 GRU 기반의 VAE로 통과시키지만, InterFusion[10]의 경우 1차원 합성곱 인코더 디코더로 구성된 필터링 과정을 거친다. InterFusion[10]의 평균 F1 점수가 0.783으로, 0.447의 성능을 보인 OmniAnomaly[7]보다 0.336 높았다. 또한 InterFusion[10]의 평균 F1 점수가 높으면서도 분산이 작은 것으로 보아 전체 데이터 집합에 대하여 안정적인 성능을 보였다. 이는 InterFusion[10]이 실제 이상 데이터에 대해서도 더욱 강건성을 가진 것으로 볼 수 있다.

또한 전반적인 VAE 성능이 AE 모델들보다 좋은 것으로 보아, VAE가 시계열 데이터의 확률적 (stochastic) 특성을 반영하기에도 유리한 것으로 해석할 수 있다.

MTAD-GAT[12]의 경우, SR 기법과 1차원 합성곱을 통해 학습 과정에서 사전 필터링을 진행한다. 그러나 GDN[14]는 사전 필터링을 거치지 않고 바로 특징 임베딩 과정을 거쳐 그래프 구조를 학습한다. 따라서 실제 이상 데이터에 대한 필터링 능력을 갖춘 MTAD-GAT[12]의 평균 F1 점수가 0.623으로 GDN[14]의 평균 F1 점수 0.411보다 0.212 높았다.

Anomaly Transformer[17]의 평균 F1 점수는 0.897로 가장 높은 평균 성능을 보였으며 MSL을 제외한 모든 데이터 집합에서도 F1 점수가 0.9 이상을 달성하였다. 특히 Table 5.에서 확인할 수 있듯이 SWaT과 WADI 데이터에서는 재현율이 1이 나왔는데, 실제 이상 데이터가 포함되었을 때도 시험 집합에 존재하는 모든 이상 데이터에 대하여 탐지를 해냈음을 보여준다.

순환 신경망의 경우, 예측을 위해 이전 시점들의 정보를 압축하는데, 수많은 정상 데이터에 의해 이상 데이터의 정보가 손실될 가능성이 있다. 그러나 Transformer의 경우 현재 시점 전까지의 정보를 압축하지 않고 입력값에 대한 모든 토큰을 병렬적으로 계산한다. 따라서 시간 정보에 대한 학습 능력이 뛰어나 AE나 VAE, GNN 모델들보다 더 좋은 성능을 보인 것으로 해석할 수 있다. 하지만 명시적으로 사전 필터링을 하지 않았기 때문에, 사전 필터링을 적용하여 더 좋은 성능을 기대해볼 수 있다.

IV. 결론 및 시사점

본 연구에서는 비지도 학습 기반의 다변량 시계열 이상 탐지 모델들을 주 역할 구조 및 사전 필터링 여부와 같은 기법에 따라 세분화하였고, 기존 이상 탐지 실험 환경과 다르게, 학습 시 잡음 또는 이상 데이터가 존재하는 현실적 데이터 집합으로 실험을 진행하였다. 실험 결과, Transformer 기반 모델을 제외하고 사전 필터링을 적용한 모델들이 적용하지 않은 모델들보다 4개의 데이터 집합에 대한 평균 F1 점수가 높았다. 이를 통해 사전 필터링이 학습 집합 내에 존재하는 잠재적인 이상 데이터 또는 잡음에 효과적임을 정량적으로 증명하였다.

특히, Anomaly Transformer 모델은 정상 데이터에 의한 지배 가능성을 고려하여 단순 순환 신경망이 아닌 기본 Transformer에 지역적 특성을 고려한 prior association을 추가하여 가장 좋은 탐지 성능을 보였다. 결과적으로 거대 신경망일수록 다변량 시계열의 표현을 효율적으로 학습하여 이상 탐지에 효과적임을 확인할 수 있었다.

본 연구를 통해 다변량 시계열 이상 탐지 과정에서 사전 필터링과 Transformer 기반 거대 신경망의 효과성을 확인함으로써, Transformer 기반의 거대 신경망에 사전 필터링 전략을 갖춰 잠재적인 이상 또는 잡음 데이터에 강건하며 탐지 성능이 좋은 모델에 관한 연구가 이루어질 것을 기대한다.

References

- [1] B. Schölkopf, J.C. Platt, J. Shawe-Taylor, A.J. Smola and R.C. Williamson, "Estimating the support of a high-dimensional distribution," *Neural computation*, vol.13 no. 7, pp. 1443-1471, Jul. 2001.
- [2] K. Choi, J. Yi, C. Park and S. Yoon, "Deep learning for anomaly detection in time-Series data: review, analysis, and guidelines," *IEEE Access*, vol. 9, pp. 120043-120065, Aug. 2021.
- [3] C. Zhang, D. Song, Y. Chen, X. Feng, C. Lumezanu, W. Cheng, J. Ni, B. Zong, H. Chen and N.V. Chawla, "A deep neural network for unsupervised

- anomaly detection and diagnosis in multivariate time series data," Proceedings of the AAAI conference on artificial intelligence, vol. 33, no. 1, pp. 1409-1416, Jul. 2019.
- [4] X. Shi, Z.Chen, H.Wang, D.Y. Yeung, W.K. Wong and W.C. Woo, "Convolutional lstm network: a machine learning approach for precipitation nowcasting," Advances in neural information processing systems, vol. 28, Dec. 2015.
- [5] J. Audibert, P. Michiardi, F. Guyard, S. Marti and M.A. Zuluaga, "Usad: unsupervised anomaly detection on multivariate time series," Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining, pp. 3395-3404, Aug. 2020.
- [6] A. Shafahi, M. Najibi, M.A. Ghiasi, Z. Xu, J. Dickerson, C. Studer, L.S. Davis, G. Taylor and T. Goldstein, "Adversarial training for free!," Advances in Neural Information Processing Systems, vol. 32, Dec. 2019.
- [7] Y. Su, Y. Zhao, C. Niu, R. Liu, W. Sun, and D. Pei, "Robust anomaly detection for multivariate time series through stochastic recurrent neural network," Proceedings of the 25th ACM SIGKDD international conference on knowledge discovery & data mining, pp. 2828-2837, Jul. 2019.
- [8] J. Chung, C. Gulcehre, K. Cho and Y. Bengio, "Empirical evaluation of gated recurrent neural networks on sequence modeling," in NIPS 2014 Workshop on Deep Learning, Dec. 2014.
- [9] D. Rezende and S. Mohamed, "Variational inference with normalizing flows," Proceedings of the 32nd International Conference on Machine Learning, vol. 37 pp. 1530-1538, Jun. 2015.
- [10] Z. Li, Y. Zhao, J. Han, Y. Su, R. Jiao, X. Wen and D. Pei, "Multivariate time series anomaly detection and interpretation using hierarchical inter-metric and temporal embedding," Proceedings of the 27th ACM SIGKDD Conference on Knowledge Discovery & Data Mining, pp. 3220-3230, Aug. 2021.
- [11] L. Dinh, J. Sohl-Dickstein and S. Bengio, "Density estimation using real nvp," NIPS 2016 Deep Learning Symposium, Dec. 2016.
- [12] H. Zhao, Y. Wang, J. Duan, C. Huang, D. Cao and Y. Tong, "Multivariate time-series anomaly detection via graph attention network," IEEE International Conference on Data Mining (ICDM), pp. 841-850, Nov. 2019.
- [13] H. Ren, B. Xu, Y. Wang, C. Yi, C. Huang, X. Kou, T. Xing, M. Yang, J. Tong and Q. Zhang, "Time-series anomaly detection service at microsoft," Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining, pp. 3009-3017, Jul. 2019.
- [14] A. Deng and B. Hooi. "Graph neural network-based anomaly detection in multivariate time series", Proceedings of the AAAI Conference on Artificial Intelligence vol. 35, no. 5, pp. 4027-4035, May. 2021.
- [15] A. Rawashdeh, M. Rawashdeh, I. Diaz and A. Ralescu, "Measures of semantic similarity of nodes in a social network." International Conference on Information Processing

- and Management of Uncertainty in Knowledge-Based Systems, vol. 443, pp. 76-85, Jul. 2014.
- [16] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A.N. Gomez, T. Kaiser and I. Polosukhin, "Attention is all you need," *Advances in neural information processing systems*, vol. 30, Dec. 2017.
- [17] J. Xu, H. Wu, J. Wang and M. Long, "Anomaly transformer: time series anomaly detection with association discrepancy," *Tenth International Conference on Learning Representations*, Apr. 2022.
- [18] A.P. Mathur and N.O. Tippenhauer, "Swat: a water treatment testbed for research and training on ICS security," *2016 International Workshop on Cyber-physical Systems for Smart Water Networks*, pp. 31-36, May. 2016.
- [19] C.M. Ahmed, V.R. Palleti and A.P. Mathur, "Wadi: a water distribution testbed for research in the design of secure cyber physical systems," *2017 International Workshop on Cyber-physical Systems for Smart Water Networks*, pp. 25-28, Apr. 2017.
- [20] D. Entekhabi, E.G. Njoku, P.E. O'Neill, K.H. Kellog, W.T. Crow, W.N. Edelstein, J.K. Entin, S.D. Goodman, T.J. Jackson, J. Johnson, J. Kimball, J.R. Piepmeier, R.D. Koster, N. Martin, K.C. McDonald, M. Moghaddam, S. Moran, R. Reichle, and J.C. Shi, M.W. Spencer, S.W. Thurman, L. Tsang and J.V. Zyl, "The soil moisture active passive (smap) mission," *Proceedings of the IEEE*, vol. 98, no.5, pp. 704-716, May. 2010.
- [21] J.P. Grotzinger, J. Crisp, A.R. Vasavada, R.C. Anderson, C.J. Baker, R. Barry, D.F. Blake, P. Conrad, K.S. Edgett, B. Ferdowski, R. Gellert, J.B. Gilbert, M. Golombek, J. Gomez-Elvira, D.M. Hassler, L. Jandura, M. Litvak, P. Mahaffy, J. Maki, M. Meyer, M.C. Malin, I. Mitrofanov, J.J. Simmonds, D. Vaniman, R.V. Welch and R.C. Wiens, "Mars science laboratory mission and science investigation," *Space Science Reviews*, vol. 170, no. 1, pp. 5 - 56, Jul. 2012.

〈 저자 소개 〉



임 주 완 (Juwan Lim) 학생회원
2022년 2월: 국민대학교 경영학과, 소프트웨어학과 학사 졸업
2022년 3월~현재: 국민대학교 컴퓨터공학과 석사과정
〈관심분야〉 Anomaly Detection, Time Series Analysis



이 재 구 (Jaekoo Lee) 정회원
2011년~2013년: LG전자 CTO부문 주임 연구원
2018년: 서울대학교 전기컴퓨터공학부 박사
2018년~현재: 국민대학교 SW융합대학 인공지능전공 조교수
〈관심분야〉 Artificial Intelligence(AI), Machine Learning(ML), Deep Learning and Data Science, Autonomous Driving, Object Detection