

다변량 시계열의 명시적 집중 학습을 통한 이상 탐지 효율성 향상

임주완¹, 노 송², 이재구^{1*}

¹국민대학교 컴퓨터공학과, ²인천대학교 정보통신공학과

*jaekoo@kookmin.ac.kr

Improved Anomaly Detection Efficiency through Explicit Attention Learning of Multivariate Time Series

Juwan Lim¹, Song Noh², Jaekoo Lee^{1*}

¹Department Of Computer Science, Kookmin University,

²Department of Information Telecommunication Engineering, Incheon University

요 약

스마트 공정 및 가상 물리 시스템(cyber physical system)의 정교함이 지속적으로 증가함에 따라 비정상적인 시스템 상태에 대한 효과적인 모니터링 및 제어에 대한 수요도 증가하고 있다. 그러나 시계열 데이터의 형태로 수집되는 방대한 양의 센서 데이터는 전문가의 분석이 필요하므로 라벨이 지정된 데이터들은 상대적으로 희소하다. 특히, 현실 세계에서는 비정상적인 상태가 거의 발생하지 않아 신경망을 훈련할 때 정상 상태 데이터와 비정상 상태 데이터의 불균형 분포를 초래한다. 최근 연구들은 이러한 한계를 보완하기 위해 트랜스포머(transformer)를 기반으로 한 비지도 학습 모델에 초점을 맞추고 있다. 그러나 트랜스포머 기반 모델들은 매개 변수의 수가 많기 때문에 많은 연산량 및 학습 시간을 필요로 하여 비효율적이다. 본 논문에서는 다변량 시계열 이상 탐지에서 트랜스포머 모델의 비효율성을 해결하기 위해, 센서 간 상관관계를 명시적으로 학습하여 연산량을 줄이고 학습 시간을 줄이는 차원 축소 접근법을 제안한다. 실험 결과 대부분의 데이터 집합에서 기존 모델 구조와 유사한 성능 수준을 달성하는 동시에 매개 변수 수를 최대 약 480 만 개에서 약 9 만 개로 감소시킬 수 있었으며, 학습에 필요한 시간을 최대 50%로 단축하였다.

I. 서 론

최근 스마트 공정 및 가상 물리 시스템(cyber physical system)이 정교해짐에 따라 비정상적인 시스템 상태에 대한 효과적인 모니터링 및 제어에 대한 수요가 증가하고 있다[1]. 특히 시스템의 각 단계에서 일정한 시간 간격으로 관측된 다양한 센서의 값을 수집하고 이에 대한 다변량 시계열 이상 탐지를 시도하는 많은 연구가 진행되었다[2].

최근 트랜스포머(transformer)[3] 기반 모델들이 대부분의 과업에서 SOTA(State-Of-The-Art)를 달성함에 따라 다변량 시계열 이상 탐지 과업에서도 이를 적용하는 시도가 있었다. 대표적으로 Anomaly Transformer[4]

는 트랜스포머의 어텐션(attention) 연산을 수행함과 동시에, 이상 데이터가 인접 시점들과 강한 상관관계를 가질 것이라는 귀납적 편향(inductive bias)인 Prior Association을 사용한 비지도 학습 모델이다. Anomaly Transformer[4]를 포함한 트랜스포머 기반 모델들은 센서 차원을 트랜스포머 차원으로 확장하여, 센서 간의 상관관계를 간접적으로 학습한다. 하지만 이러한 차원의 확장은 매개 변수의 수를 증가시켜 비효율적인 연산량의 증가와 학습 시간을 초래한다.

본 논문에서는 트랜스포머[3] 기반 모델의 차원을 센서 차원으로 축소하고, 센서 간의 어텐션 연산을 별도로 수행하여 상관관계를 명시적으로 학습하는 새로운 차원 축소 접근법을 제안한다.

실험 결과, 제안 방법은 대부분의 데이터 집합에서 기존 모델 구조와 유사한 성능 수준을 달성하는 동시에 매개 변수 수를 약 480만 개에서 데이터의 차원에 따라 최대 약 9만 개로 감소시킬 수 있었다. 또한 에폭(epoch)당 학습 시간을 모든 데이터 집합에서 50% 이상 단축하였다.

II. 본론

제안한 모델 구조는 [그림 2. 우]와 같으며 기존 모델 구조에서 센서 간의 어텐션 연산을 추가하였다. Time Series Attention에서는 센서 차원을 트랜스포머 차원으로 임베딩(embedding)하지 않고 유지하여 매개 변수 수가 방대해지는 것을 방지한다. Sensor Attention 부분에서는 MTAD-GAT[5]에서의 Feature-oriented graph Attention의 방식과 동일하게 각 센서를 노드로 설정한 완전 연결 그래프로 가정하여 그래프 어텐션 연산을 수행한다. 이때 각 노드의 임베딩 벡터는 일정 시간 간격

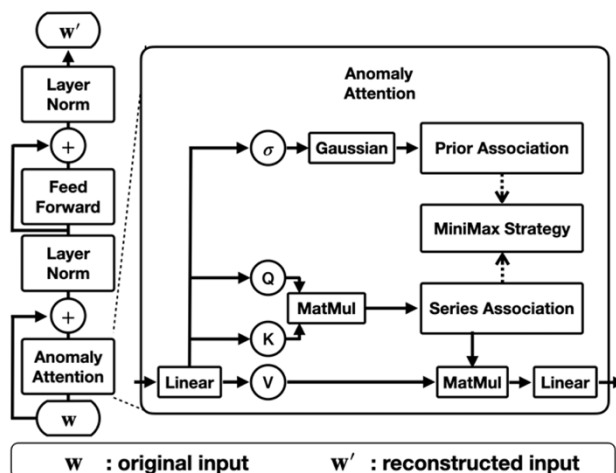


그림 1. Anomaly Transformer[4] 구조

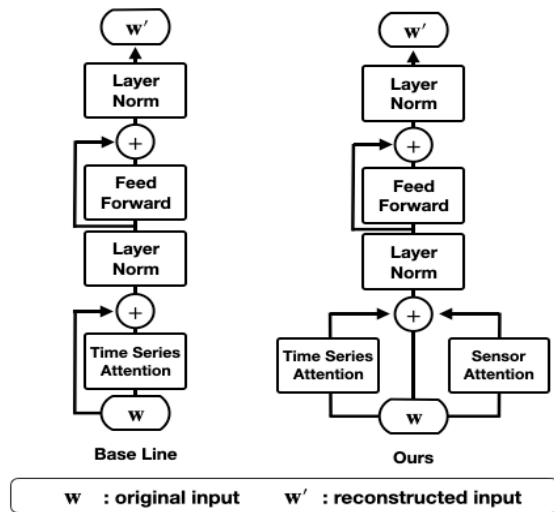


그림 1. 기본 트랜스포머 구조와 제안 방법 구조

(window size)만큼의 차원을 갖는 특징 벡터로 정의한다 [5]. 이후 원래의 입력값과 각 어텐션 연산으로부터 생성한 값을 더한 뒤 층 정규화(layer normalization)를 통과시킨다.

III. 실험

실험에 사용한 데이터 집합은 대표적인 다변량 시계열 이상 탐지 데이터 집합으로 SMD[6], MSL[7], SWaT[8]을 사용하였으며 모두 훈련 집합에 정상 데이터만을 포함하고 실험 집합에 소수의 이상 데이터가 존재한다.

모델은 기본 트랜스포머(vanilla transformer)[3]와 Anomaly Transformer[4] 두 가지를 사용하였으며, 원래의 입력값을 복원하도록 학습한다. 기본 트랜스포머의 목적함수는 평균 제곱 오차를 사용하였으며 Anomaly Transformer[4]의 목적함수는 평균 제곱 오차와 함께 Association Discrepancy[4]를 극대화하는 방향으로 학습한다.

실험 결과 (표 1)에서 볼 수 있듯이, 기본 트랜스포머에서는 MSL 기준으로 매개 변수의 수가 4,847,671개에서 89,156개로 감소하였으며 에폭 당 소요 시간도 56초에서 28초로 절반 감소하였다. 반면 (표 2)에서 볼 수 있듯이 F1 점수는 0.0726 증가하여 성능이 개선되었다.

Anomaly Transformer[4]에서는 기본 트랜스포머와는 달리 성능 수준은 기존 모델과 동일하였다. 하지만 MSL을 기준으로 매개 변수의 수가 4,859,983개에서 147,853개로 감소하였으며 학습 시간 또한 기본 트랜스포머와 동일한 감소 폭을 보였다.

IV. 결론

우리는 다변량 시계열 이상 탐지 과업을 위한 트랜스포머 기반 모델들의 효율적인 신경망 구성이 필요함을

표 1. MSL 데이터 집합에서 학습 매개변수 수 (단위 k:1,000)와 에폭 당 소요 시간(초)

| Model | Base line | | Ours | |
|-------------------------|-------------|----------------|-------------|----------------|
| | # of Params | Time per Epoch | # of Params | Time per Epoch |
| Vanilla Transformer [3] | 4,847k | 56s | 89k | 28s |
| Anomaly Transformer [4] | 4,859k | 56s | 147k | 28s |

인지하였다. 따라서 다변량 시계열의 센서 차원을 트랜스포머의 고차원으로 임베딩하지 않고 별도의 어텐션 연산을 통해 명시적으로 센서 간 상관관계를 학습하는 방법을 제안하였다. 그 결과 트랜스포머 구조를 갖는 모델들에 대해 성공적으로 성능을 개선하고 경량화하였다. 따라서 센서 차원 간 명시적 학습 방법이 트랜스포머 구조 기반 다변량 시계열 이상 탐지 모델의 성능과 경량화 수준을 향상시킬 수 있을 것으로 판단한다.

ACKNOWLEDGMENT

이 논문은 2022 년도 정부(과학기술정보통신부)의 재원으로 정보통신기획평가원의 지원을 받아 수행된 연구임(No.RS.2022-0-00516, 국가통계데이터에 적용 가능한 차등정보보호 개념을 도출하고 통계분석의 유용성을 보장해야 하는 문제 해결)

참 고 문 헌

- [1] Choi, Kukjin, et al. "Deep learning for anomaly detection in time-series data: review, analysis, and guidelines." IEEE Access, 2021
- [2] Blázquez-García, Ane, et al. "A review on outlier/anomaly detection in time series data." ACM CSUR, 2021
- [3] Vaswani, Ashish, et al. "Attention is all you need." Advances in neural information processing systems, 2017
- [4] Xu, Jiehui, et al. "Anomaly Transformer: Time Series Anomaly Detection with Association Discrepancy." ICLR, 2021.
- [5] Zhao, Hang, et al. "Multivariate time-series anomaly detection via graph attention network." ICDM, 2020.
- [6] Su, Ya, et al. "Robust anomaly detection for multivariate time series through stochastic recurrent neural network." ACM SIGKDD, 2019.
- [7] Grotzinger, John P., et al. "Mars Science Laboratory mission and science investigation." Space science reviews, 2012
- [8] Mathur, A. P., et al. "SWaT: A water treatment testbed for research and training on ICS security." 2016 international workshop on cyber-physical systems for smart water networks, 2016.

표 2. 각 데이터 집합에 대한 기존 모델과 제안 방법의 성능 비교 표

| Model | Dataset | Base Line | | | | Ours | | | |
|------------------------|---------|-----------|-----------|--------|---------------|----------|-----------|--------|---------------|
| | | Accuracy | Precision | Recall | F1 score | Accuracy | Precision | Recall | F1 score |
| Vanilla Transformer[3] | SMD | 0.9781 | 0.7843 | 0.6526 | 0.7124 | 0.9782 | 0.7870 | 0.6507 | 0.7124 |
| | MSL | 0.9546 | 0.7415 | 0.8737 | 0.8022 | 0.9738 | 0.8804 | 0.8693 | 0.8748 |
| | SWaT | 0.4252 | 0.1662 | 0.9301 | 0.2821 | 0.7590 | 0.3181 | 0.8615 | 0.4646 |
| Anomaly Transformer[4] | SMD | 0.9912 | 0.8840 | 0.9060 | 0.8949 | 0.9914 | 0.8856 | 0.9109 | 0.8981 |
| | MSL | 0.9898 | 0.9218 | 0.9866 | 0.9531 | 0.9893 | 0.9229 | 0.9807 | 0.9509 |
| | SWaT | 0.9897 | 0.9478 | 0.9684 | 0.9580 | 0.9904 | 0.9580 | 0.9754 | 0.9666 |