HAI DATASET을 이용한 최신 이상탐지 모델 성능 비교

한준서*, 양호찬*, 한윤서*, 문서진*, 정지현* 김성광*,

*화이트햇스쿨 (수료생)

Performance Comparison of Recent Anomaly Detection Models Using the HAI Dataset

Junseo Han*, Ho-chan Yang*, Yoonseo Han*, Jihyun Jeong*

*Whitehat School(Graduate student)

요 약

본 연구는 HAI 데이터셋을 활용하여 최신 이상탐지 모델들의 성능을 체계적으로 비교 분석함으로써, 기존의 LSTM 기반 모델들이 지배적인 이상탐지 연구 분야에 새로운 성과를 제시하는 것을 목적으로 한다. 현재 다양한 모델이 이상탐지 목적으로 제안되고 있지만, 실질적인 성능 평가와 비교 연구는 여전히 제한적이다. 본 연구는 최신 모델들이 실제 이상탐지 작업에서 얼마나 효과적인지 검증하고, 기존 모델 대비 성능 차이를 구체적으로 분석하여 연구의 유용성을 평가하고자 한다. 실험 결과, TranAD 모델이 F1-score 0.99, Precision 0.99, Recall 0.99라는 뛰어난 성능을 달성하였으며, 이는 기존 목표치인 0.9를 상회하는 결과로, 최신 모델의 도입이 실제 이상탐지 성능 개선에 기여할 수 있음을 확인하였다. 본 연구는 향후 이상탐지 연구에서 기존 모델을 대체할 가능성을 제시한다.

I. 서론

산업제어시스템(ICS)은 전력, 수자원 관리, 교통, 제조 등 주요 인프라를 운영하는 핵심 시스템으로, 국가 안보와 경제 안정에 중요한 역할을 한다. 그러나 인터넷 연결로 인한 사이버 공격 위협이 증가하면서, ICS 보안을 위해 비정상적인 패턴을 탐지하는 이상 탐지 시스템의 중요성이 부각되고 있다. ICS에서 생성되는 데이터는 고차원적이고 시간 순서가 있는 다변량 시간 시계열(MTS) 형태로, 기존 모델로 이상 탐지를 수행하는 데 어려움이 있다. 본 연구는 ICS 데이터에 적합한 이상 탐지 모델을 탐색하고 성능을 평가하여, ICS 보안 강화를 위한 솔루션을 제시하는 것을 목표로 한다.

II. 배경 및 관련 연구

2.1 다변량 이상탐지 접근 방법

다변량 데이터의 이상 탐지 접근법은 예측 기반과 복원 기반 두 가지로 나눌 수 있다. 예측 기반은 예측된 값과 실제 값의 편차를 측정해 이상치를 탐지하며, LSTM이 대표적이다. 복원 기반은 입력 데이터를 복원하며 이상치를 탐지하고, Autoencoder가 그 예이다. 예측 모델은 미래 값 예측에 집중하고, 복원 모델은 시계열 데이터의 전체 분포를 잘 포착하는 강점이 있다[4].

2.2 HAI 데이터셋

HAI 데이터셋은 실제 산업제어시스템(ICS) 환경에서 수집된 다양한 센서 데이터와 시스템 로그를 포함하여, 현실적인 이상 행동을 반영한 고품질의 데이터셋이다. ICS의 복잡한 환경에서 발생할 수 있는 다양한 이상 행동을 포함하고 있어, 이 데이터셋은 AI 기반 이상 탐지 모델의 성능을 평가하고 개선하는 데 적합하다.

본 연구에서는 이러한 HAI 데이터셋을 활용하여 AI 기반 이상 탐지 모델을 개발하고 성능을 평가하였다. 이를 통해, 연구 결과의 실용성

을 높이고, ICS 보안을 강화하는 데 기여하고자 한다[10].

III. 제안방법

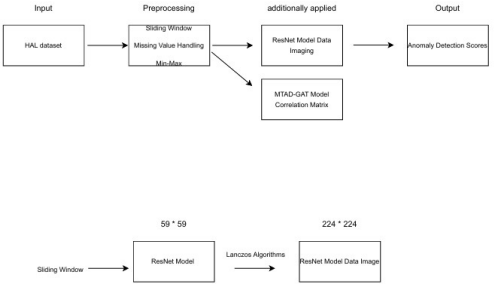

그림 1 데이터 전처리 Flow Diagram

3.1 공통 데이터 전처리

이 연구에서는 다양한 모델들이 요구하는 데이터 전처리 방법의 차이를 고려하여, 모든 모델에 공통적으로 적용되는 데이터 전처리 절차를 수립하였다. 모델별로 추가로 적용되는 전처리 방법은 별도로 기술할 예정이다.

3.1.1 슬라이딩 윈도우

시계열 데이터의 시간적 의존성을 포착하기 위해 슬라이딩 윈도우 기법을 활용한다. 고정된 크기의 시간 간격으로 데이터를 나누어, 모델이 다양한 시간 구간에서 패턴을 학습할 수 있게 하며, 윈도우 크기는 데이터와 모델 특성에 맞춰 최적화하였다.

3.2 데이터 정규화

모델의 학습 균형을 위해 Min-Max 정규화와 표준화를 적용하여 데이터의 스케일 차이를 제거하였다. Min-Max 정규화는 각 feature의 값을 [0, 1] 또는 [-1, 1] 범위로 조정하여, 특정 feature가 지나치게 큰 값이나 작은 값을 가질 때 발생할 수 있는 학습 불균형을 방지하며, 이러한 정규화는 모델 성능 향상에 기여한다.

3.3 데이터 이미지화

시계열 데이터는 슬라이딩 윈도우 기법을 통해 59x59 크기의 2차원 행렬로 변환하여 이미지화된다. 각 픽셀은 데이터값을 나타내며, 시계열 특징이 공간적 패턴으로 변환된다. Lanczos 알고리즘을 사용해 이미지를 224x224 크기로 업스케일링하여 CNN 모델의 표준 입력 크기에 맞췄으며, 이는 ResNet 모델에 사용되었다.

3.4 상관행렬 계수

다변량 시계열 데이터에서 중요한 특징을 선택하기 위해 그래프 어텐션 네트워크(GAT)를 활용할 수 있다. GAT는 각 변수를 그래프의 노드로 표현하고, 노드 간의 관계를 학습하여 변수들 간의 상관성을 파악한다. 본 연구에서는 상관 행렬을 계산하는 대신 GATConv 계층의 어텐션 메커니즘을 통해 노드 간 중요도를 학습하여 중요한 특징을 추출하고, 이를 바탕으로 모델 학습을 진행한다. 이 접근 방식은 모델의 학습 효율을 높이고, 과적합 문제를 줄이는 데 기여한다.

IV.결과 분석

Dataset	HAI 20.07		
Metric	F1	PRE	REC
ResNet	0.67	0.60	0.98
Autoencoder	0.73	0.96	0.60
TranAD	0.99	0.99	0.99
MTAD-GAT	0.95	0.97	0.94

표 1 HAI 데이터 셋 성능 평가 비교

비교 결과, TranAD 모델이 Self-conditioning, 적대적 학습, 동적 임계값 설정 기법의 조합으로 가장 우수한 성능을 보였다. 본 연구에서 비교한 ResNet, MTAD-GAT, AutoEncoder, TranAD 중 ResNet은 전처리 문제로 인해 가장 낮은 성능을 나타냈고,

MTAD-GAT는 어텐션 메커니즘을 통해 특징을 잘 학습했지만 계산 복잡도가 높았다. AutoEncoder는 레이블이 부족한 데이터에서 유용했으나, 정상 데이터를 비정상으로 오탐지할 가능성이 있었다. TranAD는 높은 성능을 기록했으나 훈련 시간이 긴 단점이 있었다.

V.결론

본 논문에서는 ICS 보안을 위해 HAI 데이터셋을 활용하여 최신 이상 탐지 모델들인 TranAD, ResNet, MTAD-GAT, AutoEncoder의 성능을 비교 분석하였다. 각 모델은 특정 상황에서 강점을 보였지만, 환경에 따라 성능에 편차가 있었다. 이러한 결과를 통해, 다양한 ICS 환경에 적합한 모델 선택이 중요함을 확인하였다. 향후 연구에서는 모델의 경량화와 성능 최적화를 통해 실제 현장에서 다양한 이상 탐지 모델이 효과적으로 활용될 수 있도록 하는 연구가 필요하다.

[참고문헌]

[1] Kim, B.; Alawami, M.A.; Kim, E.; Oh, S.; Park, J.; Kim, H. (2023). A Comparative Study of Time Series Anomaly Detection Models for Industrial Control Systems. Sensors, 23(3), 1310.

[2] Fu, T.-C. (2011). A review on time series data mining. Engineering Applications of Artificial Intelligence, 24(1), 164 - 181.

[3] Hwang, C., & Lee, T. (2021). E-SFD: Explainable Sensor Fault Detection in the ICS Anomaly Detection System. IEEE Access, 9, 140470-140486.

[4] Zhao, H., Wang, Y., Duan, J., Huang, C., Cao, D., Tong, Y., Xu, B., Bai, J., Tong, J., & Zhang, Q. (2020). Multivariate Time-series Anomaly Detection via Graph Attention Network. arXiv preprint arXiv:2009.02040.

[5] Laptev, N., Amizadeh, S., & Flint, I. (2015). Generic and scalable framework for automated time-series anomaly detection. In Proceedings of the 21st ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (pp. 1939 - 1947). ACM.

[6] Malhotra, P., Ramakrishnan, A., Anand, G., Vig, L., Agarwal, P., & Shroff, G. (2016). LSTM-based encoder-decoder for multi-sensor anomaly detection. arXiv preprint arXiv:1607.00148.

[7] Tuli, S., Casale, G., & Jennings, N. R. (2022). TranAD: Deep Transformer Networks for Anomaly Detection in Multivariate Time Series Data. arXiv preprint arXiv:2201.07284.

[8] He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep Residual Learning for Image Recognition. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR) (pp. 770-778).

[9] M. Ganesh, A. Kumar and V. Pattabiraman, "Autoencoder Based Network Anomaly Detection," 2020 IEEE International Conference on Technology, Engineering, Management for Societal impact using Marketing, Entrepreneurship and Talent (TEMSMET), Bengaluru, India, 2020, pp. 1-6, doi: 10.1109/TEMSMET51618.2020.9557464.

[10] HAI Security Dataset Manual, Version 4.0, May 2023. "HIL-based augmented ICS (HAI) security dataset." Available under CC BY-SA 4.0.