Journal of The Korea Institute of Information Security & Cryptology ISSN 1598-3986(Print)
VOL.31, NO.3, Jun. 2021 ISSN 2288-2715(Online)
https://doi.org/10.13089/JKIISC.2021.31.3.401

산업제어시스템에서 앙상블 순환신경망 모델을 이용한 비정상 탐지*

김 효 석,[1†] 김 용 민[2‡]
[1]전남대학교 정보보안협동과정 (대학원생), [2]전남대학교 전자상거래전공 (교수)

Abnormal Detection for Industrial Control Systems Using Ensemble Recurrent Neural Networks Model*

HyoSeok Kim,[1†] Yong-Min Kim[2‡]
[1]Interdisciplinary Program of Information Security, Chonnam National University (Graduate Student),
[2]Dept. of Electronic Commerce, Chonnam National University (Professor)

요 약

최근 산업제어시스템은 인터넷에 연결하지 않은 폐쇄적 상태로 운영하는 과거와 달리 원격지에서 데이터를 확인하고 시스템 유지보수를 위해서 개방적·통합적인 스마트한 환경으로 변화하고 있다. 반면에 상호연결성이 증가하는 만큼 산업제어시스템을 대상으로 사이버 공격이 증가함에 따라 산업 공정의 비정상 탐지를 위한 다양한 연구가 진행되고 있다. 산업 공정의 결정적·규칙적인 점을 고려하여 정상데이터만을 학습시킨 탐지 모델의 결과 값과 실제 값을 비교해서 비정상 여부를 판별하는 것이 적절하다고 할 수 있다. 본 논문에서는 HAI 데이터셋 20.07과 21.03을 이용하며, 순환신경망에 게이트 구조가 적용된 GRU 알고리즘으로 서로 다른 타임 스텝을 적용한 모델을 결합하여 앙상블 모델을 생성한다. 그리고 다양한 성능평가 분석을 통해 단일 모델과 앙상블 순환신경망 모델의 탐지 성능을 비교하였으며 제안하는 모델이 산업제어시스템에서 비정상 탐지하는데 더욱 적합한 것으로 확인하였다.

ABSTRACT

Recently, as cyber attacks targeting industrial control systems increase, various studies are being conducted on the detection of abnormalities in industrial processes. Considering that the industrial process is deterministic and regular, It is appropriate to determine abnormality by comparing the predicted value of the detection model from which normal data is trained and the actual value. In this paper, HAI Datasets 20.07 and 21.03 are used. In addition, an ensemble model is created by combining models that have applied different time steps to Gated Recurrent Units. Then, the detection performance of the single model and the ensemble recurrent neural networks model were compared through various performance evaluation analysis, and It was confirmed that the proposed model is more suitable for abnormal detection in industrial control systems.

Keywords: Abnormal Detection, Time Series Data, RNN, ICS, HAI Dataset

Received(04. 16. 2021), Modified(05. 03. 2021), Accepted(05. 03. 2021)

* 이 논문은 2021년도 정부(과학기술정보통신부)의 재원으로 정보통신기획평가원의 지원을 받아 수행된 연구임 (IITP-2019-0-01343, 융합보안핵심인재양성)

† 주저자, chil530@naver.com
‡ 교신저자, ymkim@chonnam.ac.kr(Corresponding author)

I. 서 론

기존의 다양한 산업이 인터넷에 연결하지 않은 폐쇄적 상태로 공정을 운영하였지만, 현재 산업들은 현장에 사용되는 장치들의 높은 신뢰성, 실시간성, 자동제어 등을 위해 네트워크에 연결된 산업제어시스템이 공정 상태를 일정하게 유지하고 있다[1]. 산업에 사용되는 장치 간 상호연결성이 증가하는 반면에 산업제어시스템을 대상으로 하는 수력 발전소 해킹(2015년, 미국), 키에프 정전사태(2016년, 우크라이나), 화학공장 해킹(2018년, 사우디아라비아) 등이 점진적으로 증가하고 있기 때문에 비정상 공정을 탐지하는 다양한 연구가 수행되고 있다[2].

산업 공정의 운전데이터는 제어신호에 의해 발생한 물리적인 장치들의 실제 운영결과이므로 비정상 상황을 인위적으로 정의하는 것이 어렵다. 그러므로 IT환경에서 주로 사용되는 이상탐지와 오용탐지로 비정상 여부를 판단하는 것이 거의 불가능하다. 따라서 현재 작업, 예정된 작업처럼 작업 순서(Task Sequence)가 존재하는 공정의 결정적·규칙적인 점을 착안하여 정상데이터만을 학습시킨 비정상 탐지 연구가 진행되고 있다[8-12,14,16,17]. 이러한 연구의 공통점은 크게 3가지이다. 첫 번째는 운전데이터를 시계열데이터의 관점으로 해석하였으며, 두 번째는 통계 및 머신러닝, 딥러닝 등을 이용하여 정상데이터만을 학습했다는 것이다. 마지막으로 탐지 모델의 예측 값과 실제 값의 차를 계산하고 특정 범위를 벗어나는 데이터를 비정상으로 탐지한 점이다.

그러나 임계값을 기준으로 정상과 비정상을 구분하는 과정에서 일부 오탐(False Positive) 문제가 발생하게 되는데, 본 논문에서는 시계열데이터의 타임 스텝(Time Step)을 변경하여 오탐 발생을 억제한 다수의 순환신경망 모델을 생성하고, 랜덤 포레스트(Random Forest)의 구조와 유사한 앙상블 순환신경망 모델을 통한 비정상 탐지를 제안하였다.

II. 관련 연구

2.1 산업제어시스템 운영구조

산업제어시스템 구성요소는 컨트롤러(Controller), HMI(Human-Machine Interface), 액추에이터(Actuator), 센서(Sensor) 등이며, 다양한 산업용 프로토콜에 의해 관리할 수 있는 구조로 되어 있다.

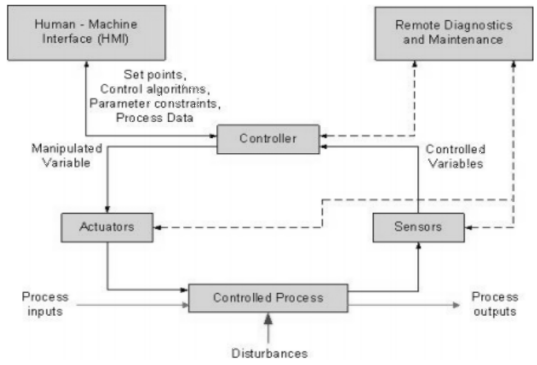

Fig. 1. Operations of ICS

Fig. 1.처럼 제어루프는 액추에이터와 센서, 컨트롤러를 사용하여 공정을 제어한다. 공정에서 발생한 속도, 온도, 압력 등을 센싱(Sensing)하여 컨트롤러에 제어 값(Controlled Variables, CV)을 전송한다. 컨트롤러는 제어 값을 해석하고 목표 값(Manipulated Variables, MV)을 생성하여 액추에이터로 보낸다. 이와 같이 모터, 스위치, 밸브 등과 같은 액추에이터는 컨트롤러의 명령에 따라 프로세스를 직접 조작하게 된다. 그리고 운영자는 HMI를 통해 공정 상태를 모니터링 및 제어하고 공정 절차에 적합한 설정 값(Set Points, SP)을 변경할 수 있는 구조이다[3].

2.2 HAI 데이터셋

일반적으로 실제 환경에서 생성된 데이터는 산업 원천기술이 포함되어 있기 때문에 외부에 공개하는 것은 보안 문제가 있으며, 실시간으로 동작하고 있는 시스템에 직접 연구를 하는 것은 가용성 보장 측면에서 어려운 부분이 존재함으로 산업제어시스템에서 비정상 탐지 연구는 테스트베드(Testbed)를 구축하여 생성된 데이터셋을 통해 진행하게 된다.

HAI(HIL-based Augmented ICS) 데이터셋은 국가보안기술연구소에서 2020년 2월 발표한 데이터셋으로 산업제어시스템 연구를 위해 GE, Emerson, Siemens 등의 산업용 제어기기, 센서, 액추에이터를 이용해 구축한 테스트베드를 기반으로 스팀-터빈, 펌프-스토리지 발전을 모방(Emulates)하여 HIL 시뮬레이터로 실제와 다름없는 산업제어시스템 환경에서 다양한 산업용 프로토콜을 OPC-UA(Open Platform Communication Unified Architecture) 표준화 프로토콜을 통해 이기종 장치의 관측 값을 수

Table 1. Comparison of HAI Dataset Versions

Release Version	Data Points /sec	Normal		Abnormal		
		Interval (hour)	Size (MB)	Attack Count	Interval (hour)	Size (MB)
HAI 20.07 (HAI 1.0)	59	177	225	38	123	181
HAI 21.03 (HAI 2.0)	78	352	471	50	112	205

집하도록 OPC-UA Gateway가 설치되었다. 데이터셋은 총 4가지(보일러, 터빈, 수처리, HIL시뮬레이터) 공정을 포함하고 있으며, 각각의 프로세스는 3가지 컨트롤러에 의해 제어되고 HIL시뮬레이터는 원격 입출력 장치를 통해 실제 프로세스와 상호 연결되는 구조이다[4]. HAI 20.07 데이터셋은 초당 59개의 Point를 수집하며, 정상데이터는 약 7일 동안 수집하고 공격데이터는 약 5일 동안 6개의 제어루프에서 38개의 공격 시나리오를 통해 수집되었다. 정상 및 공격데이터는 CSV(Comma-Separated Values)형식으로 제공되었다. 그리고 Table 1.과 같이 다양한 운영 상황과 공격시나리오를 갱신한 HAI 21.03 데이터셋이 공개되어 비교하였다[5].

또한, HAI 데이터셋을 이용 시 TaPR(Time-series Aware Precision and Recall)로 성능평가 할 것을 권장하고 있다. TaPR은 다양한 공격의 탐지여부와 탐지의 정확성과 낮은 오탐을 목표하고 있으며, 오탐 발생이 낮은 상태에서 비정상 탐지를 평가하는 TaP와 다양한 공격 범위를 찾는지 평가하는 TaR로 구성되어 2가지의 평가지표를 F1-Score로 최종 성능을 평가한다[6]. 일반적으로 정상과 비정상을 분류할 때 F1-Score를 주로 사용하지만, 산업의 운영적인 측면에서 비정상 구간을 정확히 탐지하는 것과 각각의 비정상 구간의 검출된 횟수도 측정이 가능해야 한다. 또한, 오탐에 대해서 공정을 정지하는 경우가 발생한다면 가용성 보장이 어려움으로 정밀도(Precision)와 재현율(Recall)의 평가지표에 차이가 있어야 한다. 그러므로 본 논문에서는 산업공정의 특성이 반영된 TaPR을 통해 탐지 모델의 성능을 평가한다.

2.3 공개 데이터셋을 이용한 비정상 탐지 연구

본 논문에 사용된 HAI 21.03은 최근 발표된 데이터셋으로 관련연구가 많지 않으므로 기존에 잘 알려진 공개 데이터셋을 이용한 연구를 서술하였다.

공개 데이터셋 중 SWaT(Secure Water Treatment) 데이터셋[7]을 이용한 연구에서는 정상데이터만을 학습하여 LSTM(Long Short-Term Memory), 1D CNN(Convolutional Neural Networks), GAN(Generative Adversarial Networks), Seq2Seq, Auto- Encoder 등 다양한 인공지능 모델을 사용하여 비정상 탐지하였다. 학습된 모델의 예측 값과 실제 값의 차를 계산하여 누적합(Cumulative SUM), Z-Score, GAN을 통한 판별, Auto-Encoder를 통해 생성된 재구성 데이터를 재귀하는 등 여러 가지 방법을 통해 임계값을 설정하고 비정상 탐지하였다[8-12].

SCADA network Dataset[13]을 이용한 연구에서는 매트릭스 프로파일(Matrix Profile), 계절성 아리마 모형(Seasonal Auto Regressive Integrated Moving Average, SARIMA), LSTM 총 3가지의 알고리즘을 비교분석하였으며 시계열의 유사성을 계산하거나 단기 미래 값을 예측하고 사전에 정의한 임계값을 초과하는 경우 비정상으로 예측하였다[14].

HAI 데이터셋 20.07을 이용한 연구에서는 비정상탐지를 분류문제로 해석하여 K-Nearest Neighbors, Decision Tree, Random Forest 알고리즘을 통해 성능 비교하였고[15], 정상데이터만을 학습한 연구 중 에서는 SAE(Stacked Autoencoder), SVDD(Deep Support Vector Data Description), SG-IRA(Stacked Gated Recurrent Unit-Infrequent Residual Analysis) 등의 모델을 제안하여 누적분포함수, 주파수 분석을 통해 임계값을 동적으로 학습하는 메커니즘 등을 사용하여 비정상 탐지하였다[16,17].

III. 앙상블 순환신경망을 이용한 비정상 탐지

산업제어시스템 환경에서 비정상 탐지를 할 때, 고려해야 하는 부분은 오탐 발생 여부이다. 비정상 공정이 탐지됐을 때, 현재의 물리적 공정을 일시중지하거나 현장 실사가 필요할 수 있다. 즉, 오탐에 대한 비용문제로 직결될 수 있으므로 오탐 발생을 억제한 모델을 생성하여야 한다. 그러나 오탐이 줄어들게 되면 상대적으로 미탐이 다수 발생하게 되므로 탐지 성능에 영향을 미치게 된다. 그러므로 본 장에서는 비정상 탐지 시에 시계열 기반 운전데이터의 특성으로 인해 발생할 수 있는 이슈를 살펴보고, 앙상블 순환신경망을 이용한 비정상 탐지 기법을 제안한다.

3.1 시계열 기반의 운전데이터

산업에서 발생하는 운전데이터는 운영구조에 따라 각각의 구성요소들의 시퀀스가 존재한다. 예를 들어 펌프가 가동될 때, 밸브가 잠기고 물탱크의 수위가 높아진다. 그리고 정해진 수위에 도달했을 때, 펌프의 가동이 중단된다. 그러나 구성요소들의 측정 단위나 수치의 범위가 다르기 때문에 매 초마다 데이터를 수집하는 경우에 약간의 오차가 발생할 수 있다. 또한 펌프가 가동될 때, 펌핑(Pumping)되는 속도는 빠르게 증가하지만 물탱크의 수위는 상대적으로 천천히 증가하거나 지연될 수 있으며, 물의 파동이나 수위 센서의 종류에 따라 데이터 값이 변할 수 있다. 그러므로 학습이 잘된 모델일지라도 특정 시점에서 오탐이나 미탐이 발생할 수 있다.

일반적으로 시계열데이터는 시점을 표현할 수 있는 특징들(Features), 연속되는 시점으로 구성된 기간(Time Step), 전체 시계열을 나타낼 수 있는 타임 스텝의 집합(Samples)으로 Fig. 2처럼 3차원으로 표현될 수 있다. 시계열데이터를 순환신경망에 학습할 때, 타임 스텝의 값을 너무 크거나 작게 설정한다면 적절한 예측 값을 도출하기 어렵다.

정상적인 운전데이터를 학습한 탐지 모델에 비정상데이터가 입력될 때(Fig. 2.의 4s), 3~5번째 타임 스텝까지 영향을 주게 된다. 순환신경망 모델이 현재(t)를 기준으로 다음(t+1)을 예측하는 경우, 3~5번째 타임 스텝에 비정상데이터(학습하지 않은 데이터)가 포함되어 4s~7s의 예측 값과 실제 값의 차이가 커지므로 비정상으로 탐지하게 된다.

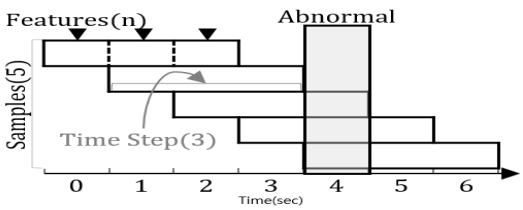

Fig. 2. 3-D Time-Series Data

3.2 순차데이터 처리에 적합한 순환신경망

비정상탐지에 사용될 모델은 데이터 종류와 특징, 그리고 사용할 기법에 따라 선택하여야 한다. 운전데이터는 산업제어시스템 구성요소의 동작에 의해 발생하고 수집된다. 산업제어시스템의 운영적인 측면에서 기계적 동작에 대한 결과가 존재하기 때문에 현재 작업과 예정된 작업이 결정적인 특성을 지니고 있다는 것을 의미한다고 할 수 있다. 그리고 시간의 흐름에 따라 물리적인 상태가 연속적으로 변화하기 때문에 각각의 구성요소들이 상호 간에 시퀀스가 존재하게 되는 순서가 있는 순차데이터이며 시계열데이터의 형태로 해석하는 것이 적절할 수 있다.

딥러닝 모델 중 순환신경망은 순차성을 기억할 수 있는 메모리 셀이 존재하기 때문에 연속된 데이터를 다루는데 적합하고 입출력데이터를 시퀀스의 길이에 상관없이 순차적으로 나타낼 수 있는 특징이 있다. 그러므로 본 논문에서는 순환신경망 중 계산복잡성을 감소시킨 셀 구조를 갖는 GRU(Gated Recurrent Units)[18]를 사용하여 학습하고 비정상 탐지 모델을 생성하였다.

3.3 제안하는 비정상 탐지 모델 구조

제안하는 앙상블 순환신경망을 이용한 비정상 탐지의 구조는 Fig. 3.과 같다. 시계열 기반의 운전데이터의 특성에 따라서 동일한 작업에서 발생되는 데이터를 측정할 때 오차가 있다는 점과 타임 스텝에 따라 예측 값이 다르게 도출되기 때문에 하나의 탐지 모델로 좋은 성능을 보장하기 어려운 부분이 있다. 또한, 비정상 공정에서 정상 공정으로 복구될 때, 탐지 모델의 타임 스텝의 크기가 큰 경우 복구된 정상 공정을 비정상으로 탐지하게 되므로 적절한 타임 스텝의 범위를 구해야 한다. 이와 같은 문제를 고려하여 앙상블 순환신경망 모델을 제안하였다.

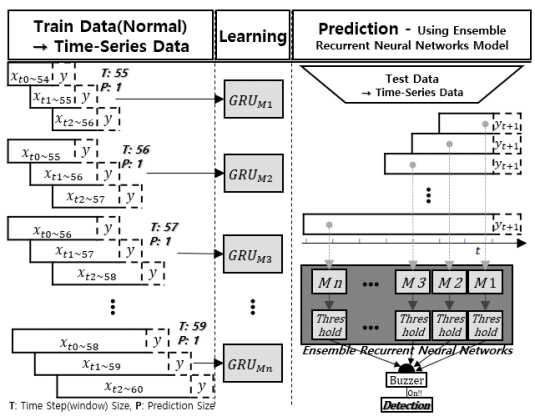

Fig. 3. Proposed Structure for Abnormal detection

제안 모델은 Fig. 3과 같이 학습에 사용되는 동일한 정상데이터를 서로 다른 타임 스텝을 갖도록 분할하고 GRU에 학습을 시켜 n개의 탐지 모델을 생성한다. 수식 1과 같이 각각의 탐지 모델은 매 번 동일한 시점을 예측($\hat{y}_{Mk}$)하고 현재 값(y)과의 차를 임계값(Threshold, T)에 의해 정상(0) 또는 비정상(1)을 판단하게 된다. 다수의 탐지 모델(Mk) 중 1개라도 비정상으로 탐지하는 경우에 현재의 공정을 비정상으로 판단(Buzzer, B 〉 0) 하였다.

$$B = \sum_{k=1}^{n} T(\hat{y}_{Mk} - y) \quad (1)$$

3.4 앙상블 순환신경망 모델의 생성 과정

모델의 생성 과정은 Fig. 4.와 같다. 데이터 전처리 과정에서 정상데이터만을 사용하였으며 최소-최대 정규화(Min-Max Normalization)하여 0~1 사이의 값으로 변환한다. 그리고 학습데이터의 NaNs (Not a Number)값 처리와 평균, 표준편차, 최소/최대 값 등을 확인하고 운전데이터에 관여되지 않은 특징을 미리 제거한 뒤 기준 모델을 생성한다.

기준 모델을 구성할 때, 시계열데이터의 타임 스텝을 결정하는 파라미터와 GRU 알고리즘의 하이퍼파라미터의 튜닝이 필요하다. 사용할 파라미터 값의 범위를 정의하고 그리드 서치(Grid Search)를 통해 경계 값 분석을 진행하여 기준 모델을 생성한다.

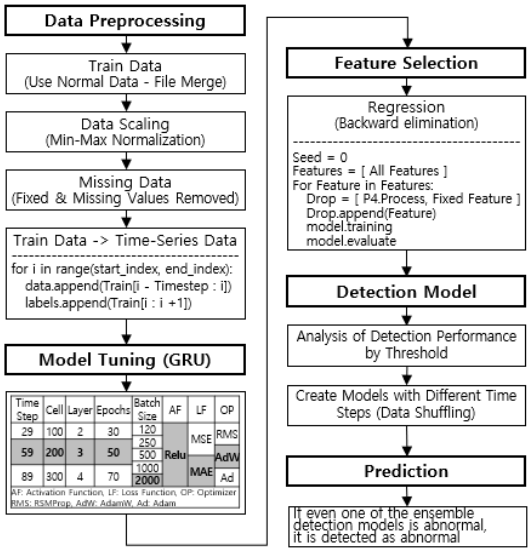

Fig. 4. Flow of Detection Model Generation

탐색적 자료 분석(EDA) 과정이 생략된 만큼 기준 모델이 좋은 성능을 갖기 어려울 수 있으므로 각각의 특징들이 기준 모델에 어떤 영향을 주는지 확인하기 위해 후진제거법(Backward Elimination)[19]을 사용하여 불필요한 특징을 제거하고 탐지 모델을 생성한다. 그 다음 타임 스텝마다 모델을 생성할 때, 학습데이터는 셔플(Shuffle)하여 서로 다른 타임 스텝을 갖는 모델들을 생성한다. 그 외의 파라미터 등의 설정 값은 모두 동일한 조건에서 진행한다. 타임 스텝의 길이와 동일한 학습데이터를 통한 학습순서를 변경하면 모델의 예측오차가 미세하게 달라진다. 즉, 하나의 모델에서 발견하지 못한 비정상을 탐지가 가능하다. 이때, 손실률과 오차 값을 분석하여 오탐이 발생하지 않도록 모델 별 임계값을 설정하고 테스트 데이터를 모델에 입력하여 예측하도록 한다.

IV. 실험 및 분석

본 장에서는 서로 다른 타임 스텝을 갖는 모델을 통해 앙상블 순환신경망 모델의 성능을 도출하였다.

4.1 실험 환경

실험은 HAI 데이터셋(HAI 20.07, HAI 21.03)을 이용하였고 학습에는 Colab(Google Colaboratory)을 사용하였다. Colab은 브라우저를 통해 Python 코드를 작성하고 실행할 수 있으므로 머신러닝, 데이터 분석 및 교육에 적합하다. Colab Pro는 T4, P100 GPU와 최대 25.51GB RAM을 사용 가능하기 때문에 Colab Pro를 통해 실험하였다[20].

4.2 탐지 모델 생성

GRU를 이용한 탐지모델 생성과정에서 고려한 사항은 크게 3가지이다. 첫 번째는 운전데이터는 정상데이터만을 학습하는 것이며, 공정에서 측정된 특징들은 범위가 다르기 때문에 스케일링하여 3차원 형태의 시계열데이터로 변환하였다.

두 번째는 탐지모델의 하이퍼파라미터를 튜닝하는 과정이다. Table 2.처럼 각 파라미터의 범위를 설정하여 가장 성능이 좋은 값을 선택하였다. 앙상블 순환신경망을 구성할 때 시계열데이터를 무작위로 학습하는 것과 타임 스텝의 길이를 다르게 적용한 앙상블 모델을 생성하기 위해 55~59의 수치를 선택하였다.

Table 2. Parameters Used for Model

No	Parameter	Range	Select
1	Time Step	29 ~ 119	**55 ~ 59**
2	RNN	LSTM, GRU	Stacked GRU
3	Cell(Node)	100 ~ 300	200
4	Hidden Layer	2 ~ 4	3
5	Epoch	30 ~ 80	50
6	Batch Size	250 ~ 2000	2000
7	Activation Function	Relu	Relu
8	Loss Function	MSE, MAE	MAE
9	Optimizer	RMSProp, Adam, AdamW	AdamW[21]
10	Dropout	0.1 ~ 0.3	X
11	Data Shuffle	True, False	**True**

세 번째는 특징선택이다. 특징추출 및 선택기법은 다양하지만 새로운 특징을 추출하는 경우에는 학습데이터의 다양성, 사이즈 등에 따라서 기존의 특징을 설명할 수 있는 새로운 변수가 매번 달라질 수 있으므로 특징선택 기법 중 불필요한 특징들을 제거하는 후진제거법을 사용하였다. 먼저, 운전 간 발생하는 특징만을 이용하기 위해 HIL시뮬레이터(P4)와 데이터에 변화가 없는 특징을 제거하여 HAI 20.07에서 59개의 특징 중 11개, HAI 21.03에서 78개의 특징 중 28개의 특징을 제거하였다. 후진제거법을 사용하여 제거할 특징을 선택하는 기준은 최종 모델의 임계값 설정을 고려하여 정상데이터의 예측오차가 낮게 측정되면서 모든 시점의 예측오차가 일정하도록 유지하기 위해 1회에는 모델의 손실률과 학습데이터의 절대오차의 최대 값, 평균절대오차 등을 통해 최초 기준모델을 생성하고, 2~3회에서는 특징을 제거했을 때 전반적인 성능이 향상되면서 TaP가 높은(오탐이 낮은)상태에서 TaR이 높은(미탐이 낮은) 경우의 특징을 찾는 것을 목표로 TaP에 비중을 두고 가장 좋은 성능이 확인되는 특징을 제거하였다.

HAI 20.07의 경우, Table 3.과 같이 P1.보일러(B2004, B4002, FCV02D, FT01Z, PCV02Z), P2.터빈(24Vdc, P2_VT01e)에서 총 7개의 특징을 제거하고, 학습한 모델은 Table 4.와 같이 임계값을 0.07로 설정하였을 때 TaPR를 기준으로 87.3%의 성능을 보이고 있으며 38개 공격시나리오 중 34개를 탐지하였다. 미탐지된 시나리오는 11번(P3_SP01, P3_LCP01), 12번(P3_SP02, P3_LCV01), 22번(P3_SP01, P3_LCP01), 25번(P1_B2016)이다.

HAI 21.03의 경우, Table 5.와 같이 P1.보일러(FCV02D, PIT01), P2.터빈(SIT01, VXT02, VYT02)에서 총 5개의 특징을 제거하고, 학습한 모델은 Table 6.와 같이 임계값을 0.04로 설정하였을 때 TaPR을 기준으로 93.3%의 성능을 보이고 있으며, 50개 공격시나리오 중 48개를 탐지하였다. 미탐지된 시나리오는 10번(P1_B2016), 25번(P3_LCV01D)이다.

Table 3. Feature Selection using Backward Elimination in HAI 20.07 (Remove checked features)

	1Round							2Round				3Round			
Drop	Loss	Train Data		Test Data			Drop	Test Data			Drop	Test Data			Drop
Features	MAE	Max	Mean	F1	TaP	TaR		F1	TaP	TaR		F1	TaP	TaR	
P1_B2004	2.034	0.068	0.007	0.117	0.063	0.789		0.828	0.949	0.734		0.873	0.980	0.787	V
P1_B4002	2.108	0.070	0.007	0.211	0.122	0.765		0.873	0.973	0.791	V	-	-	-	V
P1_FCV02D	1.817	0.051	0.007	0.279	0.169	0.803	V	-	-	-	-	-	-	-	V
P1_FT01Z	2.034	0.067	0.007	0.573	0.455	0.776	V	-	-	-	-	-	-	-	V
P1_PCV02Z	2.008	0.068	0.007	0.836	0.920	0.766	V	-	-	-	-	-	-	-	V
P2_24Vdc	1.457	0.065	0.005	0.191	0.107	0.843	V	-	-	-	-	-	-	-	V
P2_VT01e	1.809	0.061	0.007	0.236	0.138	0.821	V	-	-	-	-	-	-	-	V
Result	1.288	0.040	0.005	**0.851**	**0.902**	**0.805**		**0.873**	**0.973**	**0.791**		**0.873**	**0.980**	**0.787**	

Table 4. Evaluation of Detection Performance by Threshold in HAI 20.07

	eTaPR				Classification Evaluation Metrics			
Threshold	F1	TaP	TaR	Detected	Accuracy	Precision	Recall	F1 score
0.065	0.871	0.963	0.796	34/38	0.984	0.873	0.711	0.784
0.070	0.873	0.980	0.787	34/38	0.984	0.877	0.696	0.776
0.090	**0.831**	**0.981**	**0.721**	**32/38**	**0.982**	**0.892**	**0.633**	**0.741**
0.095	0.828	**0.987**	0.713	32/38	0.982	**0.893**	0.618	0.731

Table 5. Feature Selection using Backward Elimination in HAI 21.03 (Remove checked features)

	1Round							2Round				3Round			
Drop	Loss	Train Data		Test Data			Drop	Test Data			Drop	Test Data			Drop
Features	MAE	Max	Mean	F1	TaP	TaR		F1	TaP	TaR		F1	TaP	TaR	
P1_FCV02D	3.644	0.061	0.008	0.925	0.978	0.878	V	-	-	-	-	-	-	-	V
P1_PIT01	3.735	0.060	0.008	0.914	0.965	0.868		0.929	0.982	0.883	V	-	-	-	V
P2_SIT01	3.529	0.057	0.008	0.921	0.980	0.868	V	-	-	-	-	-	-	-	V
P2_VXT02	3.590	0.054	0.008	0.919	0.976	0.869	V	-	-	-	-	-	-	-	V
P2_VYT02	3.592	0.055	0.008	0.919	0.974	0.870	V	-	-	-	-	-	-	-	V
Result	3.407	0.042	0.075	**0.917**	**0.980**	**0.862**		**0.929**	**0.982**	**0.883**		**0.929**	**0.982**	**0.883**	

Table 6. Evaluation of Detection Performance by Threshold in HAI 21.03

	eTaPR				Classification Evaluation Metrics			
Threshold	F1	TaP	TaR	Detected	Accuracy	Precision	Recall	F1 score
0.040	0.933	**0.975**	0.894	48/50	0.989	**0.760**	0.736	0.748
0.045	**0.929**	**0.982**	**0.883**	**48/50**	**0.989**	**0.764**	**0.711**	**0.736**
0.050	0.915	**0.979**	0.858	47/50	0.988	**0.767**	0.685	0.723
0.055	0.908	**0.982**	0.844	47/50	0.988	**0.768**	0.657	0.708

Classification Evaluation Metrics의 정밀도(Precision)를 근거하여 실제 환경에서 위의 임계값은 오탐 발생의 여지가 있기 때문에 임계값을 0.09(HAI 20.07), 0.045(HAI 21.03) 정도의 수준을 기준으로 진행하였다.

4.3 앙상블 순환신경망 모델의 성능 평가

4.3.1 기준 모델의 성능평가

앙상블 순환신경망 모델에 사용할 특징과 파라미터 등이 결정되었으므로 학습데이터는 랜덤하게 셔플링하여 55~59의 타임스텝을 갖는 5개의 모델을 생성한다. 각 모델의 성능은 Table 7.과 같다. 각각의 탐지모델들은 오탐이 발생하지 않도록 TaP의 성능이 98%의 수준을 보이며, 탐지성능은 88.8%(HAI 20.07), 92.9%(HAI 21.03)가 나타났다.

Table 7. Performance of Each Model

DATA SET	Time Step	eTaPR			
		F1	TaP	TaR	Detect
HAI 20.07	55	0.871	0.990	0.777	34/38
	56	0.867	0.981	0.777	34/38
	57	0.879	0.990	0.790	35/38
	58	0.876	0.985	0.788	34/38
	59	0.888	0.980	0.811	35/38
HAI 21.03	55	0.923	0.980	0.872	47/50
	56	0.921	0.978	0.870	47/50
	57	0.929	0.981	0.882	48/50
	58	0.923	0.982	0.871	47/50
	59	0.929	0.982	0.883	48/50

4.3.2 앙상블 순환신경망 모델의 성능평가

생성된 각 모델을 모두 결합한 제안 모델의 분류 평가지표에 대한 성능은 Table 8.과 같다. 정확도는 데이터셋의 버전에 따라 98.6%, 98.9%로 확인됐으나 HAI 20.07를 기준으로 정상데이터 427,073개, 공격데이터 17,527개로 클래스가 불균형하기 때문에 모델의 성능 해석에 어려움이 있다. 또한 정밀도, 재현율, F1-Score는 테스트데이터의 크기와 공격시나리오의 수에 따라서 차이가 발생하고 시계열데이터의 구조상 공격이 끝난 후 공정이 정상적으로 동작하기 위해 복구하는 과정을 비정상으로 탐지하여 오탐이 발생한 것으로 평가되어 실제 환경에서 평가지표로 사용하기 어려운 부분이 있다. 그러므로 Table 9.의 eTaPR 지표가 보다 적절한 것으로 나타났다.

Table 8. Classification Evaluation Metrics of the Proposed Model

DATASET	Classification Evaluation Metrics			
	Accuracy	F1-Score	Precision	Recall
HAI 20.07	0.986	0.810	0.873	0.755
HAI 21.03	0.989	0.751	0.755	0.747

Table 9. eTaPR of the Proposed Model

DATASET	eTaPR			
	F1	TaP	TaR	Detect
HAI 20.07	**0.898**	0.974	0.833	35/38
HAI 21.03	**0.938**	0.977	0.902	48/50

다음은 Table 9.의 eTaPR을 기준으로 설명하였다. HAI 20.07을 이용한 실험에서 오탐의 발생여부와 관련이 있는 TaP는 97.4%로 나타났다. 실험 간 발생한 오탐은 특정 구간에서 13초(10/29 16시 00분 17초~29초), 1초(10/31 19시 30분 19초) 발생하였고, 그 외에는 공격 이후의 안정화되지 않은 공정을 탐지한 것으로 실제 오탐 발생량이 낮은 것을 확인하였다. 총 38개의 공격 시나리오 중 3개(11번, 12번, 22번)의 공격시나리오에서 미탐이 확인되었으며 5개의 모델을 사용한 최종 성능은 TaPR을 기준으로 89.8%로 나타났다.

HAI 21.03을 이용한 실험에서 발생한 오탐은 공격 후의 안정화되지 않은 구간을 탐지한 것 외에 오탐 발생은 없었으며, 총 50개의 공격 시나리오 중 2개(10번, 25번)을 미탐하였다. 5개의 모델을 사용한 최종 성능은 TaPR을 기준으로 93.8%로 나타났다.

4.3.3 기준모델과 제안모델의 ROC Curve 비교

Fig. 5.의 AUC - ROC Curve 결과는 앙상블을 형성하는 기준모델들의 오탐이 0에 가까운 수치에서 83~85%의 성능을 보이고 있으며, 제안모델은 오탐

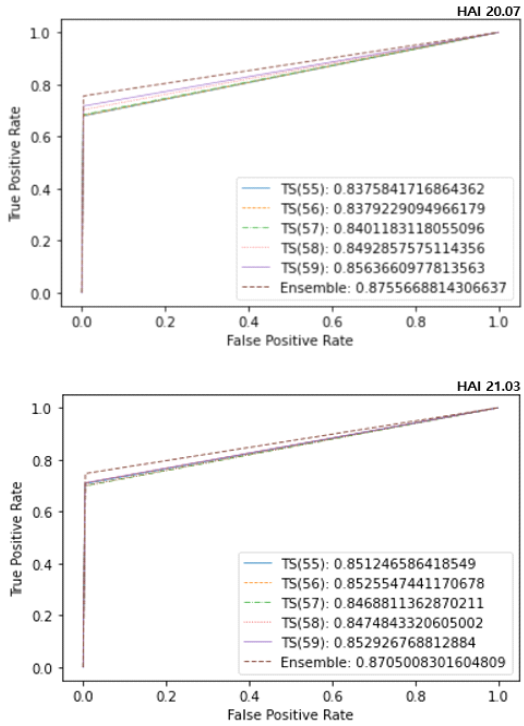

Fig. 5. AUC - ROC Curve

발생이 0에 가까운 수치에서 2% 증가한 87%의 성능을 보이기 때문에 각 모델들이 미탐에 대해 상호보완적 작용을 하고 있다고 설명할 수 있다. 기준 모델을 늘려간다면 탐지 성능이 증가할 것으로 보인다.

V. 결론 및 향후 연구

본 논문에서는 산업제어시스템에서 발생하는 비정상 공정을 앙상블 순환신경망을 이용하여 비정상 탐지를 제안하였으며, HAI 20.07의 정상데이터만을 학습하여 비정상 탐지한 기존 연구[16,17] 중에서 SAE모델의 정확도/ROC Curve(0.9603/0.6646), SG-IRA의 F1-Score/TaP/TaR(81.1%/93.4%/71.1%)과 비교하였을 때, 제안모델의 성능이 보다 높은 것으로 나타났다.

산업제어시스템 환경에서는 오탐이 발생하지 않도록 하면서 전체적인 탐지성능을 높여야 하는 문제를 고려해야 한다. 운영구조에 따라 특정 구성요소에 문제가 발생한다면 해당 공정 전체에 연쇄적인 영향을 주기 때문에 후진제거법을 통해 특징을 선택하고 GRU를 사용하여 기준 모델을 생성하였다. 그리고 시계열의 타임 스텝과 학습데이터를 랜덤하게 셔플링하여 동일 데이터를 다른 관점에서 학습하고, 오탐을 고려하여 정밀도를 향상시킨 다수의 순환신경망 모델을 앙상블하게 구축하여 탐지성능을 향상시킬 수 있음을 확인하였다.

그러나 학습데이터 및 특징 부족으로 인해 특정 공정의 비정상 탐지가 어려운 부분이 있어 비정상을 탐지할 수 있는 새로운 특징을 추출하거나 물리적인 공정에 구성요소를 추가하여야 할 것으로 나타났다. 또한, 전체 공정에 대한 비정상 여부를 탐지하였기 때문에 비정상 탐지 후 대응과정에서 비정상 공정을 유발시킨 장치를 특정 지을 수 있어야 보다 빠른 대응 및 복구가 가능할 것으로 추가적인 연구를 수행할 예정이다.

References

[1] Choi Seungoh and Kim Woo-Nyon, "Cyber-Physical System TestBed Technology Research Trend," Jonornal of The Korea Institute of information Security & Cryptology, 27(2), pp. 46-56, Apr. 2017

[2] Kwon Sungmoon and Shon Taeshik, "SWaT testbed dataset and Abnormal Detection Trend," Jonornal of The Korea Institute of information Security & Cryptology, 29(2), pp. 29-35, Apr. 2019

[3] Keith Stouffer, Victoria Pillitteri, Suzanne Lightman, Marshall Abrams and Adam Hahn, Guide to Industrial Control System Security, NIST SP 800-82, May. 2015

[4] Hyuk-ki Shin, Womyo Le, Jeong-Han Yun, and Hyoungchun Kim, "HAI 1.0: HIL-based Augmented ICS Security Dataset," 13th USENIX Workshop on Cyber Security Experimentation and Test, 2020

[5] ICS(Industrial Control System) Security Dataset·Github, "HIL-based Augmented ICS(HAI) Security Dataset", https://github.com/icsdataset/hai (last accessed 01-Apr-2021)

[6] Won-seok Hwang, Jeong-Han Yun, Jonguk Kim, and Hyoungchun Kim, "Time-Series Aware Precision and Recall for Anomaly Detection - Considering Variety of Detection Result and Adresing Ambiguous Labeling," CIKM'19: Procedings of the 28th ACM International Conference on Information and Knowledge Management, 2019

[7] iTrust Labs_Dataset Info - iTrust, "Secure Water Treatment (SWaT)", https://itrust.sutd.edu.sg/itrust-labs_datasets/dataset_info/ (last accessed 15-Nov-2020)

[8] Jonathan Goh, Sridhar Adepu, Marcus Tan and Lee Zi Shan, "Anomaly Detection in Cyber Physical Systems Using Recurrent Neural Networks," 2017 IEEE 18th International Symposium on High Assurance Systems Engineering (HASE), Jan, 2017

[9] Moshe Kravchik and Asaf Shabtai, "Detecting Cyber Attacks in Industrial Control Systems Using Convolutional Neural Networks," Proceedings of the 2018 Workshop on Cyber-Physical Systems Security and PrivaCy, Dec, 2018

[10] Simon Duque Anton, Lia Ahrens, Daniel Fraunholz and Hans Dieter Schotten, "Time is of the Essence: Machine Learning-Based Intrusion Detection in Industrial Time Series Data," 2018 IEEE International Conference on Data Mining Workshops (ICDMW), Nov, 2018

[11] Dan Li, Dacheng Chen, Jonathan Goh and See-kiong Ng, "Anomaly Detection with Generative Adversarial Networks for Multivariate Time Series," International Workshop on Big Data, Streams and Heterogeneous Source Mining: Algorithms, Systems, Programming Models and Applications, Jan, 2019

[12] Jonguk Kim, Jeong-Han Yun and Hyoung Chun Kim, "Anomaly Detection for Industrial Control Systems Using Sequence-to-Sequence Neural Networks," Workshop on the Security of Industrial Control Systems & of Cyber-Physical Systems(CyberICPS 2019) in conjunction with ESORICS 2019, Nov, 2019

[13] antoine-lemay·Github, "SCADA network datasets", https://github.com/antoine-lemay/Modbus_dataset (last accessed 15-Nov-2020)

[14] Julien Audibert, Pietro Michiardi, Frédéric Guyard, Sébastien Marti and Maria A Zuluaga, "USAD: UnSupervised Anomaly Detection on Multivariate Time Series," Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining, Aug. 2020.

[15] Mokhtari S, Abbaspour A, Yen KK and Sargolzaei A, "A Machine Learning Approach for Anomaly Detection in

Industrial Control Systems Based on Measurement Data," Electronics. 10(4), Feb. 2021
〔16〕 Doyeon Kim, Chanwoong Hwang, and Taejin Lee, "Stacked-Autoencoder Based Anomaly Detection with Industrial Control System," Software Engineering, Artificial Intelligence, Networking and Parallel/Distributed Computing. SNPD 2021. Studies in Computational Intelligence, vol. 951, pp. 181-191, Feb. 2021
〔17〕 Xingchao Bian, "Detecting Anomalies in Time-Series Data using Unsupervised Learning and Analysis on Infrequent," Journal of Institute of Korean Electrical and Electronics Engineers, 24(4), pp. 1011-1016, Dec. 2020
〔18〕 Kyunghyun Cho, Bart van Merrienboer, Caglar Gulcehre, Dzmitry Bahdanau, Fethi Bougares, Holger Schwenk and Yoshua Bengio, "Learning Phrase Representations using RNN Encoder-Decoder for Statistical Machine Translation," In Proceedings of the Empirical Methods in Natural Language Processing (EMNLP), Sep, 2014
〔19〕 Nahyeon Ryu, Hyungseok Kim and Pilsung Kang, "Evaluating Variable Selection Techniques for Multivariate Linear Regression," Journal of the Korean Institute of Industrial Engineers, 42(5), pp. 314-326, Oct, 2016
〔20〕 Google Colaboratory, "Colab Pro", https://colab.research.google.com/notebooks/pro.ipynb (last accessed 01-Apr-2021)
〔21〕 Ilya Loshchilov and Frank Hutter, "Decoupled weight decay regularization," In International Conference on Learning Representations(ICLR), Jan, 2019

〈저자소개〉

김 효 석 (HyoSeok Kim) 정회원
2018년 8월: 전남대학교 대학원 정보보안협동과정 석사
2018년 9월~2020년 8월: 전남대학교 대학원 정보보안협동과정 박사과정
2016년 2월~현재: 안랩 근무
〈관심분야〉 침입탐지 및 대응, 융합보안, 인공지능

김 용 민 (Yong-Min Kim) 종신회원
2002년 8월: 전남대학교 전산통계학과 박사
2006년~현재: 전남대학교 문화콘텐츠학부 전자상거래전공 교수
전남대학교 대학원 정보보안협동과정 교수
〈관심분야〉 시스템 및 네트워크 보안, 전자상거래 보안, IoT 융합보안 등