산업제어시스템의 운전 환경에서 앙상블 순환 신경망을 이용한 비정상 공정 탐지

전남대학교 대학원
정보보안협동과정
박사과정 김효석

목차

1-1. 연구 배경 및 목적

- 사이버 물리 시스템을 통해 현실세계의 사람·센서·액추에이터 등을 인터넷에 연결하여 제어할 수 있는 환경으로 전환
- 운영 환경의 변화는 공격 접점(Attack Surface)을 확대함으로 산업제어시스템을 대상으로 하는 사이버 위협의 증가
- 물리적 공정을 연결하는 구간에서 수집가능한 운전 데이터를 통해 비정상 공정을 탐지하여 운영 가용성·안전성 보장

연도	대상	피해 내용
2010	원자력(이란)	스턱스넷(Stuxnet) 악성코드를 통해 유지보수PC에 침투, 원자력 제어시스템을 조작하여 원심분리기 파괴.
2011	송유도(미국)	스카다 시스템의 ID, 비밀번호를 탈취하고 원격접속을 통해 제어 컴프 시설을 공격하여 가동 중단.
2012	정유사(사우디아라비아)	샤문(Shamoon) 악성코드 공격으로 내부 PC 3만 대 감염 및 네트워크 마비.
2014	제철소(독일)	제어시스템의 파로로 통로로 제어 실능.
2015	전력망(우크라이나)	블랙에너지(Black Energy) 악성코드를 통해 제어시스템을 공격하여 6개 주요 도시 22만여 가구 정전 발생.
2016	원자력(독일)	원전 근무자가 USB 사용 중 악성코드 감염으로 운영 중지.
2017	자동차(일본)	워너크라이 랜섬웨어(WannaCry Ransomware)에 감염 되어 2일 동안 생산 및 조립 중단.
2017	비상전이원(미국)	내부망 해킹으로 달러스의 비상 사이렌이 15시간 동안 가동.
2017	화학공장(사우디아라비아)	트리톤(TRITON) 공격으로 석유화학공장 가동 중단.
2018	반도체 생산(대만)	생산설비 업그레이드 중 USB를 통한 악성코드 유입으로 공장 가동 중단, 손실 초정액은 110억 원.
2019	알루미늄 생산(노르웨이)	랜섬웨어에 감염되어 유럽, 미국 등의 공정이 비정상적으로 폭주하기 9개월 이상의 시간이 걸린 후에도 정상상태로 복구하지 못함.
2019	발전소(태국수력원)	수력발전소 설비가 사이버 공격을 받으면서 중단(Shutdown), 19개 주의 전력공급이 차단.
2020	자동차(일본)	Ekans 악성코드 감염으로 사내 네트워크로 침투하여 생산라인 관리시스템 장애, 차량 출하 중단.

1-2. 연구 내용 및 범위

- 산업제어시스템 보안 및 비정상 탐지 연구 조사
- 공개 데이터셋 비교를 통한 데이터 수집
- 산업제어시스템 운영 및 구성요소에 따른 운전 데이터 해석
- 인공지능 모델 학습을 통한 문제 해결 방법 선정(회귀, 예측 모델)
- 산업제어시스템 환경을 고려한 비정상 탐지 모델 생성
- 탐지 모델에서 추출한 예측오차를 통한 비정상 공정 도출

1-1. 산업제어시스템 참조 아키텍처 및 구성요소

- 물리적 설비는 Level 0 위치(엑추에이터, 센서 등)
- 물리적 설비의 논리제어장치(PLC 등)는 Level 1 위치
- 운전데이터는 Level 1~2에서 해석

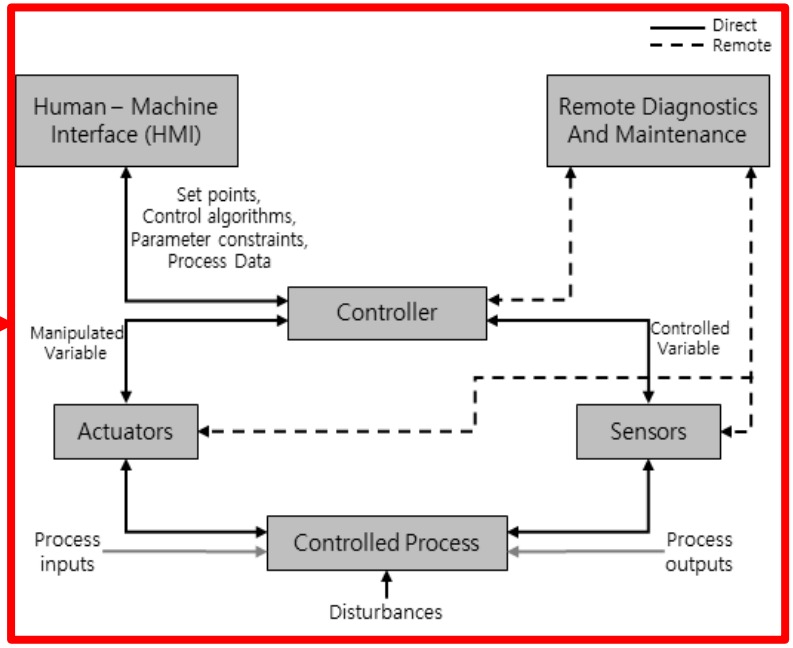
  (OPC-UA Gateway, PI System 등)

- 산업용 프로토콜사용

  (Modbus, S/comm, Ethernet/IP(Industrial Protocol), …)

- 순환 명령 구조 (

  센서 측정(+운전자 명령) -> 컨트롤러 해석 / 명령 -> 프로세스 조작)



1-2. 산업제어시스템 보안 요구사항

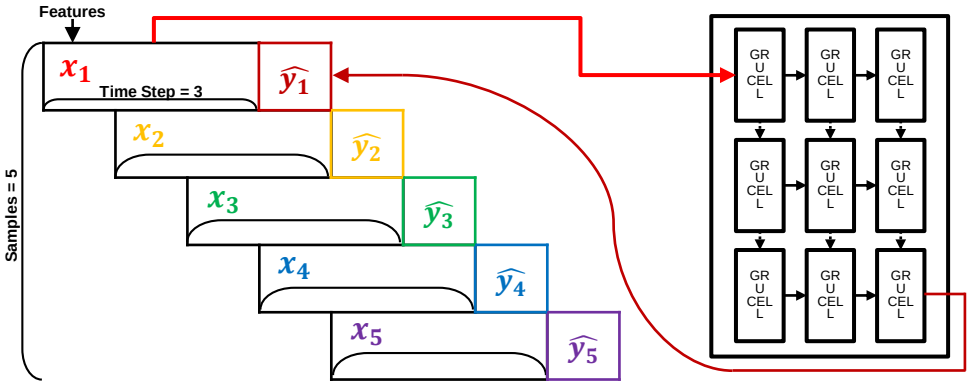
- KS표준의 보안요구사항 중 시스템 무결성(“통신 무결성 “, “결정론적 출력 “)과 이벤트적시 대응(“지속적인 모니터링 “)에 해당하며, 정상 작업의 연속성을 유지 및 보장하여야 함.

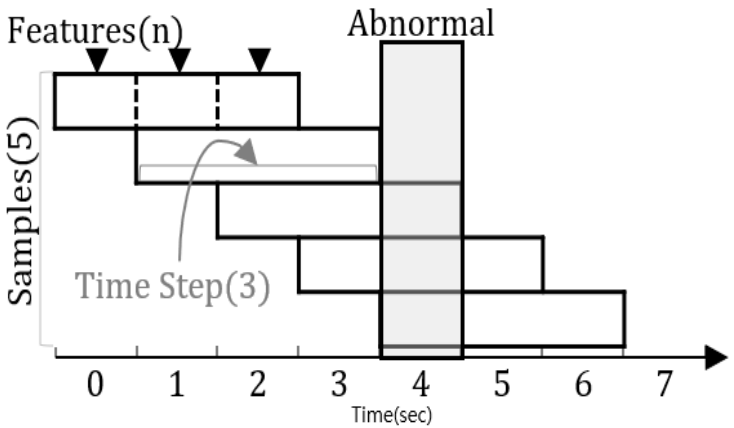
기본 요구사항	설명	근거	컴포넌트요구사항
식별 및 인증	시스템 또는 자산에 대한 접근을 허용하기 전에, 모든 사용자를 식별하고 인증한다.	인가 받지 않은 사용자가 컴포넌트 접근 자격을 획득하는 것을 방지하기 위해 필요.	인간사용자 식별 및 인증, 소프트웨어 및 프로세스와 장치 식별 및 인증, 식별 관리, 식별자 관리, 인증자 관리, 무선 접근 관리, 애플리케이션 보안 인증 강도, PKI 인증, PKI 인증 강도, 인증자 피드백, 모호한 시도 실패, 시스템 사용 알림, 신뢰할 수 없는 네트워크의 영유 접근, 대칭키 인증 강도(14개)
사용통제	인증된 사용자의 활동을 권한을 기초로 수행하여 컴포넌트에 허용된 작업을 수행, 권한의 사용을 모니터링 한다.	사용자의 작업 수행을 허용하기 전에 필요한 권한이 부여되었는지 검증하여 인가되지 않은 작업으로부터 보호.	권한 부여, 무선 사용 통제, 휴대용 및 모바일 장치에 대한 사용 통제, 모바일 코드, 세션 잠금, 원격 세션 종료, 동시 세션 제어, 감사 이벤트, 감사 저장 용량, 감사 처리 실패 시 응답, 모바일 스탠드, 무인 장치, 물리적 진단 및 시험 인터페이스의 사용(13개)
시스템 무결성	인가 되지 않은 조작 또는 변경을 방지하기 위한 컴포넌트의 무결성을 보장한다.	물리적 자산의 무결성은 설계 실행 중이거나 저장소에 저장 중 또는 유지보수 절차 중과 같이 온line 또는 온line 중이지 않을 상태 모두에서 유지되어야 한다.	통신 무결성, 암호로 보호, 보안 기능 검증, 소프트웨어 및 정보 무결성, 입력 검증, 결정론적 출력, 오류처리, 세션 무결성, 감사 정보 보호, 업데이트되지 않은, 물리적 탬퍼 인지 및 경보, 제품 공급자의 신뢰기반(RoT) 루트, 자산 소유자 신뢰기반(RoT) 루트, 무도난 모바일 무결성(14개)
데이터 기밀성	인가 받지 않은 공개를 방지하기 위해 저장소에 저장된 데이터와 통신세션 정보의 기밀성을 보장한다.	권한 컴포넌트가 설정한 정보는 기밀 또는 민감한 특성이 있으므로 인가 받지 않은 접근으로부터 보호되어야 한다.	정보 기밀성, 정보 지속성, 암호 사용(3개)
데이터 흐름제한	제어 시스템을 구역(zone) 및 통신경로(conduit)로 분할하여 허용되지 않은 데이터 흐름을 제어한다.	제어시스템 네트워크를 공용 네트워크로부터 분리하는 것부터 단일방 파이어월이, 상태기반 방화벽, 또는 DMZ 구성 사용에 이르기까지 다양한 아키텍처를 포함	네트워크 분할, 구역 경계 보호, 익명 피어 투 피어(P2P) 통신 제한, 애플리케이션 분할(4개)
이벤트적시 대응	보안사고 발견 시 적절한 담당자에게 통보하고, 필요한 경우 증거를 보존하며, 적시에 시정조치를 취하여 보안 위반에 대응한다.	안전한 상태유지를 보장하기 위해 시스템을 모니터링하여야하며, 이벤트가 시스템의 보안에 영향을 주는 경우, 적시 통보가 관련된 위험을 완화하기 위해 중요한	감사 보호 접근성, 지속적인 모니터링(2개)
자원 가용성	필요 서비스에 대한 서두름 없는 저하에 대응하여 컴포넌트의 가용성을 보장한다.	이 컴포넌트 요구사항 시리즈의 목표는 컴포넌트가 필요한 유형의 서비스 거부 이벤트에 대응, 가용성 보장해야 한다	서비스 거부(DoS) 방지, 자원 관리, 제어시스템 백업, 제어시스템 복구 및 재구성, 비상 전원, 네트워크 및 보안 구성 설정, 최소 기능, 제어시스템 컴포넌트 복귀(8개)

2-1. 시계열 데이터와 순환 신경망

- 운영 구조에 따라 모든 Point는 현재 값은 다음 값에 영향을 주고, 시간에 흐름에 따라서 변한다.

- 순환 신경망은 시퀀스의 길이에 상관없이 유연한 구조이며, 연속적인 데이터 처리에 적합하다.





(x_1)

```
'input'  : tensor([[9.9751e-01, 5.7956e-01, 4.3588e-01, ..., 8.6484e-01, 8.8776e-01, 2.5674e-04],
                   [9.9751e-01, 5.7918e-01, 4.3588e-01, ..., 8.6795e-01, 8.8535e-01, 2.5674e-04],
                   [9.9751e-01, 5.8562e-01, 4.3588e-01, ..., 8.7054e-01, 8.8396e-01, 2.5674e-04],
                   ...,
                   [9.9751e-01, 5.6099e-01, 4.3588e-01, ..., 8.7520e-01, 7.5754e-01, 2.5674e-04],
                   [9.9751e-01, 5.6043e-01, 4.3588e-01, ..., 8.7281e-01, 7.5583e-01, 2.5674e-04],
                   [9.9751e-01, 5.5416e-01, 4.3588e-01, ..., 8.7026e-01, 7.5406e-01, 2.5674e-04]])
```

$(\widehat{y_1})$

```
'predict': tensor([9.9744e-01, 5.5319e-01, 4.3593e-01, 4.1077e-01, 5.2651e-01, 9.9953e-01,
                   9.9383e-01, 5.6975e-01, 9.9968e-01, 9.9966e-01, 2.7198e-04, 1.6308e-01,
                   1.6238e-01, 3.9501e-01, 6.1492e-01, 9.8606e-01, 9.9624e-01, 3.7718e-01,
                   4.2834e-01, 2.6622e-01, 3.2905e-01, 5.0664e-01, 2.8673e-01, 2.8645e-01,
                   1.8055e+00, 4.7633e-02, 9.6405e-02, 2.8687e-02, 4.3906e-01, 6.1533e-01,
                   5.1694e-01, 3.8916e-01, 7.2084e-01, 3.5193e-01, 3.4857e-01, 5.8513e-03,
                   1.5504e-03, 8.6702e-01, 7.5447e-01, -3.9041e-05])
```

2-2. 앙상블 학습 및 모델 결합

- Voting
  - 각 모델의 결과(0,1,0,2,0,1)를 다수결 투표를 통해 최종 결과값(0)을 도출한다.
- Bagging
  - 중복을 허용한 상태로 랜덤 샘플링한 데이터를 학습(Bootstrap), 학습 결과를 평균하여 최종 결과 (Aggregation) 를 도출
- Boosting
  - 선행 모델의 결과 중 잘못 분류된 데이터에 높은 가중치를 주어 후행 모델에서 재 학습하여 오류를 개선
- Stacking
  - 다양한 이기종 모델에서 얻은 결과를 다음 모델의 학습 데이터로 사용하고 최종 모델에서 예측
- Buzzer
  - 본 논문에서 사용된 “Buzzer” 는 학습데이터 순서를 랜덤으로 모두 학습하며 서로 다른 타임스텝을 갖는 내부모델을 생성한다. 내부 모델은 오탐을 억제한 상태이며, 미탐지에 대해 상호보완함

- 좋은 모델을 찾는것 보다, 높은 탐지 성능을 도출하는것이 목표
  - 우승자를 찾는 것이 아닌, 되도록이면 다 같이 모든 문제를 맞출 수 있도록 함

- 제안 모델은 비정상 상황에 누구든 하나라도 탐지하면 비정상으로 간주하여 성능을 높임
  - 오탐이 없으면, 상대적으로 미탐 발생률이 높을 수 있으므로 해당 내용을 개선.
  - 오탐이 억제된 모델의 특성 상, 비정상으로 탐지 시 정확도가 아주 높음.

출처: http://www.koreaittimes.com/news/articleView.html?idxno=63654

3-1. 산업제어시스템 데이터셋

No	Dataset	Type	Procotol	Data	
				Format	Anonymization
1)	SecureWater Treatment(SWaT)	Water Treatment	Modbus/Ethernet/IP	csv	X
2)	ICS Cyber Attack Datasets	Power System, Gas Pipeline, Water, EMS	Modbus	csv, arff	△
3)	SCADA Network Datasets	SCADA	Modbus	csv,pcap	X
4)	HAI Dataset	Power System (Boiler, Turbine, Water)	OPC-UA (Modbus, S7, ···)	csv	X

- HAI Dataset을 제외한 기존 데이터셋의 문제
  - 시간 동기화
  - 비정상 상태의 라벨링
  - 환경에 따른 공격의 다양성

Release Version	Data Points /sec	Normal		Abnormal		
		Interval (hour)	Size (MB)	Attack Count	Interval (hour)	Size (MB)
HAI2.0/ (HAI1.0)	59	177	225	38	123	181
HAI21.03 (HAI20)	78	362	471	50	112	206

4-1. 산업제어시스템 비정상 탐지 연구 동향

비정상 탐지 모델의 성능 비교 연구

연구자	년도	비교 모델	데이터셋	주요내용
J. Inoue et al.	2017	DNN(LSTM), one-class SVM	SWaT	SVM은 정상 범위를 벗어나는 값을 탐지하는 데 효과적이나 오탐지를 보고하는 경우가 있으며, 전반적으로 DNN의 성능이 좋음.
SDD Anton et al.	2018	SVM, Random Forest, K-NN, K-means clustering	SCADA network	여러 모델을 사용하여 비정상 탐지에 적합한 모델을 찾아 성능 비교하였으며, 최적의 SVM과 Random Forest가 서로 다른 데이터셋에서 도출되었음.
SDD Anton et al.	2018	Matrix Profile, SARIMA, LSTM	SCADA network	매트릭스 프로파일은 계절성 패턴의 변형은 전반적으로 적절하게 탐지하였으나, LSTM은 학습시간이 오래 걸리는 데이터의 특성에 크게 의존하는 부분이 있으며 다른 알고리즘과 하이퍼파라미터의 적절한 튜닝이 필요함.
G. Bernieri et al.	2019	SVM, Random Forest, K-NN, One-class SVM, Autoencoder	SWaT	Random Forest와 K-NN은 탐지 정확도가 높게 측정되어, 비지도 학습보다 지도 학습 기반의 알고리즘이 우수함 것으로 나타남.
S Mokhtari et al.	2021	K-NN, Decision Tree, Random Forest	HAI	정상과 비정상 데이터의 불균형 문제를 SMOTE를 통해 적절한 문제를 해결한 후 학습하였을 때, Random Forest가 가장 좋은 성능을 보임.
D Kim et al.	2021	SAE, SVDD	HAI	SAE(Stacked Autoencoder), SVDD(Deep Support Vector Data Description)를 비교하였을 때, SAE결과가 SVDD에 비해 좋은 성능을 보임.

비정상 탐지 기법 연구

연구자	년도	모델·기법	데이터셋	주요내용
J. Goh et al.	2017	LSTM-RNN, 누적합	SWaT	비정상 여부에 대한 판단은 예측값과 실제값의 차이가 특정 임계치의 초과여부로 판별 함.
M. Kravchik et al.	2018	CNN, Z-Score	SWaT	데이터 전처에 걸친 예측값과 실제값의 차를 계산한 후 각 예측 및 실제 값 사이의 절대 값을 도출하고, 각 특징을 정규화 하여 예측 오류의 가능성에 대한 Z-Score를 통해 비정상 탐지 함.
D. Li et al.	2019	LSTM-RNN, GAN	SWaT	정상 학습 조건에서 LSTM을 사용하여 다변량 시계열 분포를 파악하고, GAN을 통해 새 주성 데이터와 실제 데이터의 차이를 판별하여 비정상 탐지 함.
J. Kim et al.	2019	Seq-to-Seq	SWaT	정상 데이터셋을 학습하여 탐지 단계에서 이전에 관찰된 값을 기반으로 예측값을 도출하고 모델이 예측한 값과 실제값의 차이가 비정상 탐지의 기준으로 사용 함.
Y Hu et al.	2019	CNN, 네트워크 공통 특징 비교 연결	KDD 99, SCADA network	IT 네트워크의 특징은 KDD CUP 99 데이터셋에서 ICS 네트워크의 특징은 SCADA 데이터셋에서 추출하여, 두 네트워크의 흐름 데이터에 대한 공통된 특징을 비교하고 연결하여 각 네트워크 간의 정보 관계를 증명 함.
Xingchao Bian	2021	GRU, 주파수 분석	HAI	GRU를 통한 예측을 통해 비정상 탐지하며, 주파수 분석을 통해 실제값을 동적으로 학습하는 메커니즘을 사용 함.
Y.G Kim et al.	2021	SOMAD, 예측오류 에러	SWaT, HAI	SOM(Self-Organizing Map)을 이용하여 예측오류 패턴을 학습하고 일련의 오류 패턴의 궤적을 추정하여 기설점검을 통해 비정상을 탐지.

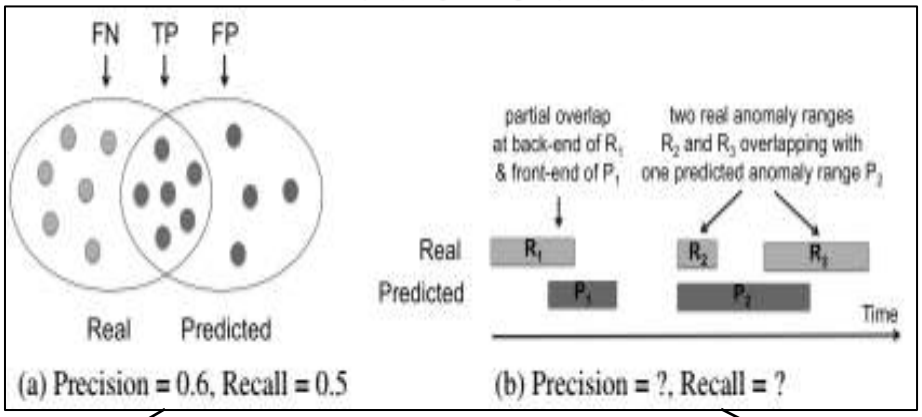
4-2. 선행 연구 고찰 및 제한점

- 선행 연구 고찰
  - 학습 방법: 지도 학습, 비지도 학습
  - 문제 정의: 분류, 회귀 문제
  - 탐지 기법: 예측오차, 누적오차, 정상데이터의 일반오차 등의 경계(임계 값) 설정
  - 성능 평가: 분류성능평가지표를 사용 함

- 제한점
  - 이기종 모델의 탐지 모델 및 탐지 기법에 따라 비정상 탐지 내용이 상이함(단일 모델의 한계)
  - 비정상 탐지의 타겟이 전체 공정으로 제한되어 있음
  - 성능 평가 시, 오탐지에 대한 언급이 부족함(현실성을 고려하여 오탐 발생량 및 기간을 분석해야 함)
  - 분류 모델의 경우, 새로운 유형의 공격을 탐지하기 어려움
  - 회귀 모델의 경우, 회귀 모델(+산업 환경)에 적합한 성능평가지표를 사용해야 함

4-3. 비정상 탐지성능 평가지표

분류 / 회귀 성능평가의 차이

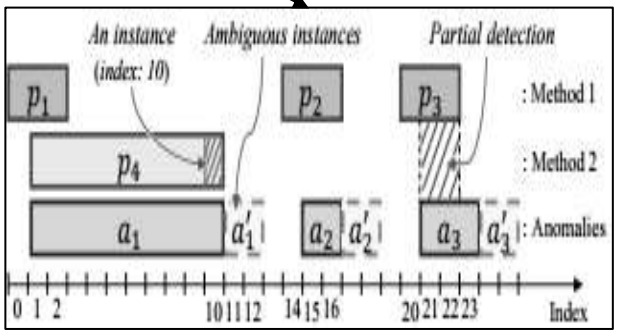


Point Based

Confusion Matrix		Actual	
		Positive	Negative
Predicted (Detected)	Positive (비정상)	Ture Positive	False Positive
	Negative (정상)	False Negative	Ture Negative

Classification Evaluation Metrics

Range Based



TaPR(Time-series Aware Precision and Recall)

1-1. 비정상 탐지 연구에 대한 고려사항

- 실험 데이터는 HAI 20.07, 21.03이며, 성능평가지표는 TaPR을 사용한다.

  HAI 21.03을 기준으로 작성하며, HAI 20.07은 특징 선택 및 앙상블 구성 시에 HAI 21.03과 비교를 위해 실험한다.

  또한, 분류 성능평가지표를 일부 포함한다.

- 데이터 구조는 시계열 데이터를 기반으로 한다.

- 정상 데이터만을 학습한다.

- 학습에 사용되는 특징은 운전 데이터의 포인트(Point)를 변경하지 않거나 본래의 데이터로 복원이 가능하도록 한다.

- 산업제어시스템의 구성요소 변경 및 교환 등을 고려하여 모든 포인트를 포함하여 비지도 학습 기반 회귀 예측 모델을 사용한다.

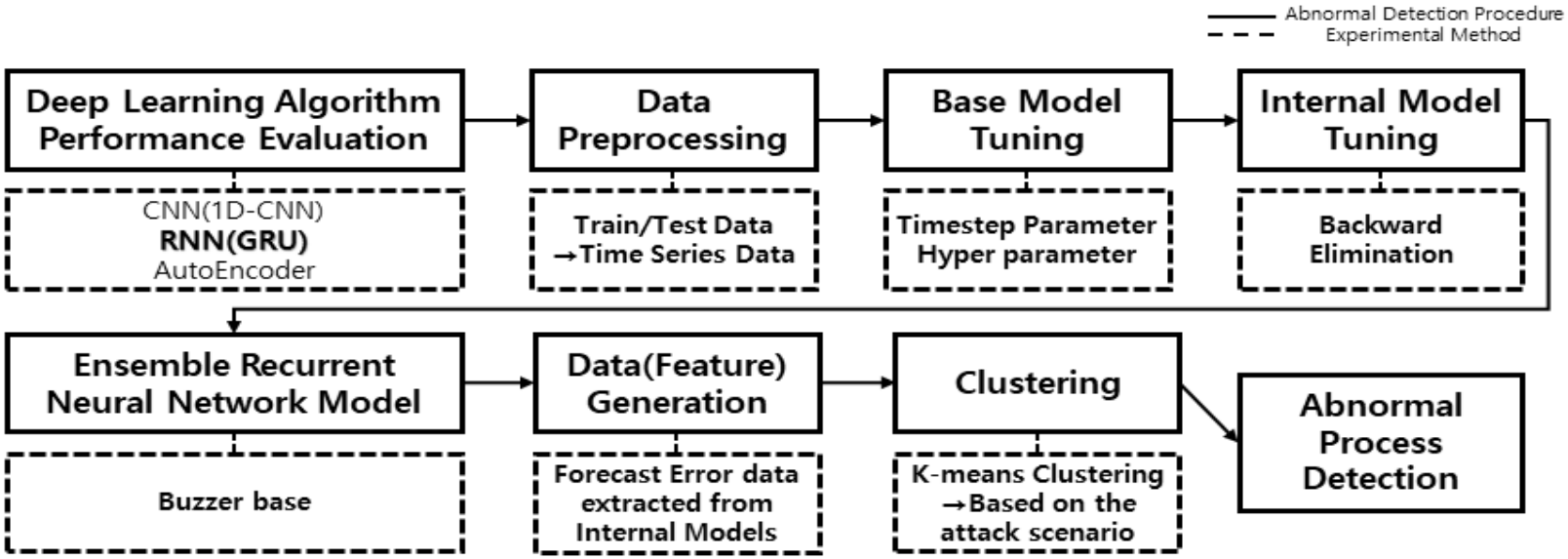
- 학습된 모델의 예측값과 실제값의 차이를 평균하여 정상과 비정상을 식별한다.

- 탐지 모델은 되도록 오탐이 발생하지 않도록 임계값을 조정한다.

- 비정상 탐지 후 앙상블 모델을 구성하는 내부 모델들의 평균예측오차를 통해 비정상 탐지를 유발시킨 공정을 도출한다.

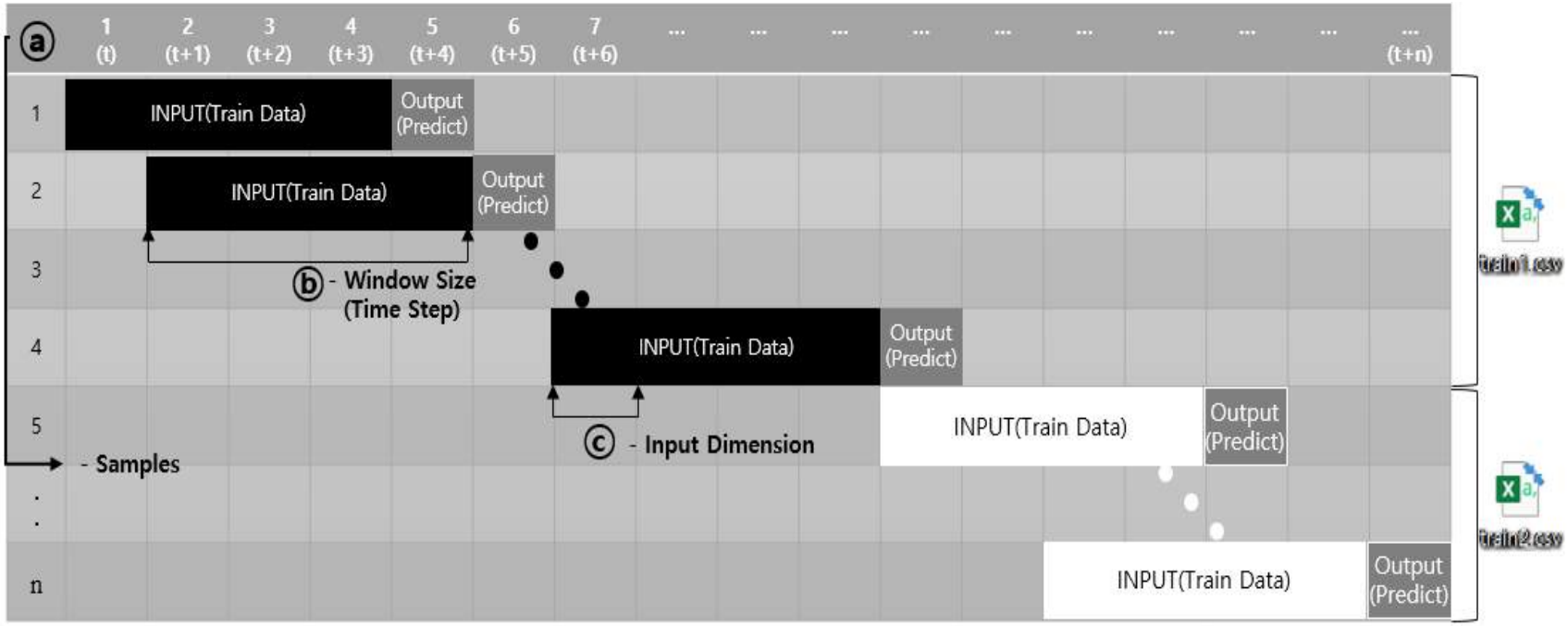
2-1. 비정상 공정 탐지 전체 흐름도

- 학습 데이터 전처리
- 순환 신경망을 기반으로 하는 기초, 내부, 앙상블 모델 학습 후 비정상 탐지
- 앙상블 순환 신경망 모델의 예측오차 데이터 생성
- 공격 시나리오 기준으로 평균예측오차 데이터를 K-평균 클러스터링하여 비정상 공정 도출



3-1-1. 학습데이터 전처리

- Null/NaNs 값 확인
- 데이터 스케일링
  - Min-Max Normalization
- 시계열 데이터로 변환



3-1-2. 기초 모델 생성(모델 튜닝)

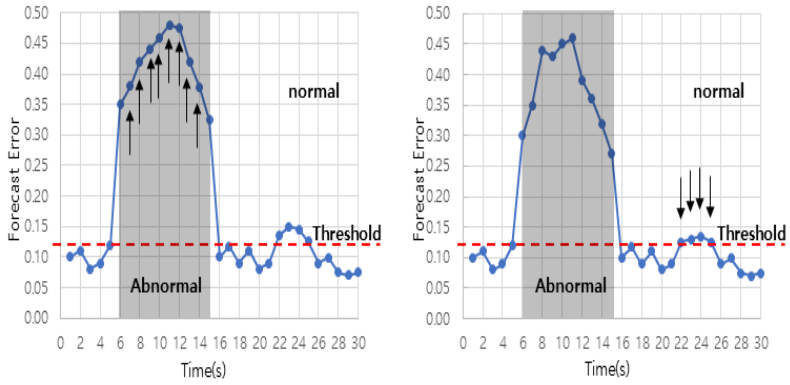
- 파라미터
  - TimeStep
- 하이퍼파라미터
  - Cell, Layer, Epoch, …

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
10	Dropout	0.1 ~ 0.3	X
11	Data Shuffle	True, False	True

3-1-3. 내부 모델 생성(특징 선택)

- 후진 제거법(Backward Elimination): 전체 특징을 포함시킨 상태에서 독립변수(설명변수)를 하나씩 제거하여 설명력이 가장 작은 변수를 제거.

- 검증시, 특정 특징을 제거한 상태에서 학습데이터를 다시 모델에 입력하여 최대 오차 값과 평균 오차가 낮게 나타나면 설명력이 가장 작은 변수로 판단.



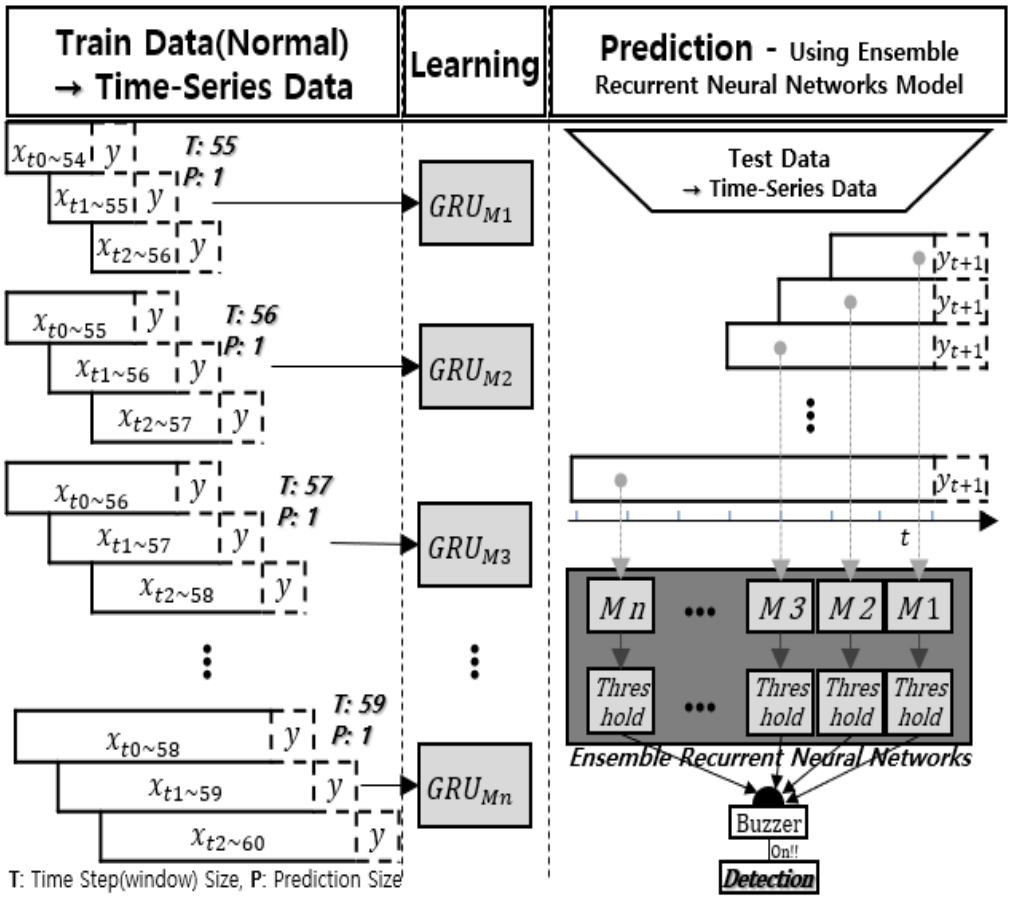
	1Round							2Round				3Round			
Drop(D)	Loss	Train Data		Test Data			D	Test Data			D	Test Data			D
Features	MAE	Max	Mean	F1	TaP	TaR		F1	TaP	TaR		F1	TaP	TaR	
P1_B2004	3.754	0.062	0.008	0.918	0.978	0.864		0.925	0.979	0.876		0.919	0.975	0.868	
P1_B2016	3.707	0.061	0.008	0.915	0.978	0.860		0.909	0.947	0.874		0.916	0.949	0.886	
P1_B3004	3.749	0.061	0.008	0.909	0.971	0.854		0.916	0.976	0.862		0.911	0.977	0.853	
P1_B3005	3.749	0.062	0.008	0.899	0.976	0.833		0.907	0.973	0.850		0.907	0.971	0.851	
P1_B4002	3.836	0.061	0.008	0.918	0.974	0.868		0.921	0.971	0.875		0.916	0.960	0.876	
P1_B4005	3.736	0.061	0.008	0.920	0.976	0.869		0.920	0.966	0.878		0.919	0.975	0.870	
P1_B400B	3.705	0.061	0.008	0.922	0.969	0.880		0.925	0.978	0.878		0.934	0.976	0.896	
P1_B4022	3.713	0.062	0.008	0.909	0.956	0.866		0.924	0.975	0.878		0.923	0.972	0.878	
P1_FCV01D	3.768	0.061	0.008	0.920	0.977	0.869		0.927	0.978	0.881		0.921	0.971	0.876	
P1_FCV01Z	3.719	0.061	0.008	0.921	0.977	0.871		0.922	0.968	0.880		0.912	0.955	0.872	
P1_FCV02D	3.644	0.061	0.008	0.925	0.978	0.878	V	-	-	-	-	-	-	-	V
P1_FCV02Z	3.760	0.062	0.008	0.911	0.967	0.862		0.922	0.971	0.878		0.919	0.952	0.889	
P1_FCV03D	3.725	0.062	0.008	0.887	0.968	0.819		0.901	0.956	0.852		0.923	0.972	0.879	
P1_FCV03Z	3.726	0.061	0.008	0.907	0.973	0.849		0.920	0.978	0.869		0.924	0.981	0.874	
P1_FT01	3.739	0.061	0.008	0.909	0.952	0.869		0.912	0.950	0.877		0.914	0.957	0.875	
P1_FT01Z	3.701	0.061	0.008	0.909	0.965	0.859		0.923	0.981	0.871		0.927	0.964	0.892	
P1_FT02	3.723	0.061	0.008	0.914	0.961	0.872		0.921	0.974	0.874		0.930	0.972	0.891	
P1_FT02Z	3.739	0.060	0.008	0.919	0.974	0.869		0.923	0.975	0.876		0.924	0.976	0.877	
P1_FT03	3.725	0.061	0.008	0.906	0.959	0.859		0.919	0.970	0.874		0.923	0.978	0.874	
P1_FT03Z	3.726	0.061	0.008	0.900	0.962	0.846		0.917	0.969	0.871		0.924	0.978	0.876	
P1_LCV01D	3.716	0.060	0.008	0.904	0.979	0.840		0.903	0.964	0.850		0.914	0.982	0.856	
P1_LCV01Z	3.733	0.061	0.008	0.918	0.976	0.867		0.909	0.958	0.865		0.915	0.962	0.872	
P1_LIT01	3.730	0.061	0.008	0.915	0.979	0.859		0.925	0.986	0.872		0.929	0.986	0.878	
P1_PCV01D	3.716	0.061	0.008	0.914	0.975	0.859		0.921	0.970	0.877		0.920	0.977	0.869	
P1_PCV01Z	3.710	0.061	0.008	0.913	0.978	0.857		0.915	0.976	0.862		0.907	0.953	0.865	
P1_PCV02Z	3.683	0.061	0.008	0.913	0.975	0.858		0.921	0.965	0.881		0.920	0.973	0.872	
P1_PIT01	3.735	0.060	0.008	0.914	0.965	0.868		0.929	0.982	0.883	V	-	-	-	V
P1_PIT02	3.699	0.062	0.008	0.919	0.977	0.867		0.933	0.979	0.890		0.928	0.970	0.890	
P1_TIT01	3.706	0.061	0.008	0.919	0.971	0.873		0.914	0.958	0.874		0.913	0.959	0.872	
P1_TIT02	3.706	0.061	0.008	0.910	0.961	0.864		0.924	0.975	0.878		0.930	0.973	0.889	
P2_24Vdc	3.030	0.060	0.007	0.902	0.956	0.854		0.925	0.976	0.879		0.920	0.969	0.877	
P2_CO_rpm	3.500	0.060	0.008	0.879	0.959	0.811		0.875	0.959	0.805		0.885	0.975	0.811	
P2_HILout	3.295	0.060	0.007	0.915	0.974	0.864		0.926	0.980	0.878		0.921	0.967	0.880	
P2_SIT01	3.529	0.057	0.008	0.921	0.980	0.868	V	-	-	-	-	-	-	-	V
P2_SIT02	3.543	0.057	0.008	0.914	0.975	0.860		0.902	0.959	0.851		0.925	0.980	0.876	
P2_VT01	3.509	0.055	0.008	0.913	0.978	0.856		0.913	0.954	0.875		0.919	0.972	0.871	
P2_VXT02	3.590	0.054	0.008	0.919	0.976	0.869	V	-	-	-	-	-	-	-	V
P2_VXT03	3.632	0.052	0.008	0.918	0.982	0.862		0.930	0.977	0.886		0.925	0.974	0.880	
P2_VYT02	3.592	0.055	0.008	0.919	0.974	0.870	V	-	-	-	-	-	-	-	V
P2_VYT03	3.650	0.051	0.008	0.905	0.963	0.854		0.904	0.925	0.884		0.931	0.978	0.887	
P3_FIT01	3.668	0.062	0.008	0.914	0.973	0.863		0.924	0.973	0.880		0.927	0.978	0.880	
P3_LCP01D	3.709	0.063	0.008	0.910	0.955	0.869		0.930	0.983	0.882		0.925	0.975	0.880	
P3_LCV01D	3.697	0.061	0.008	0.912	0.961	0.867		0.917	0.976	0.865		0.921	0.975	0.872	
P3_LIT01	3.727	0.062	0.008	0.917	0.973	0.867		0.920	0.962	0.882		0.928	0.980	0.881	
P3_PIT01	3.702	0.065	0.008	0.913	0.971	0.862		0.914	0.956	0.876		0.930	0.982	0.883	
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**	

3-1-4. 앙상블 모델 생성 및 탐지

- 기초 모델을 통해 사용할 파라미터의 범위를 정하고, 내부 모델을 통해 불필요한 특징을 제거한 상태.

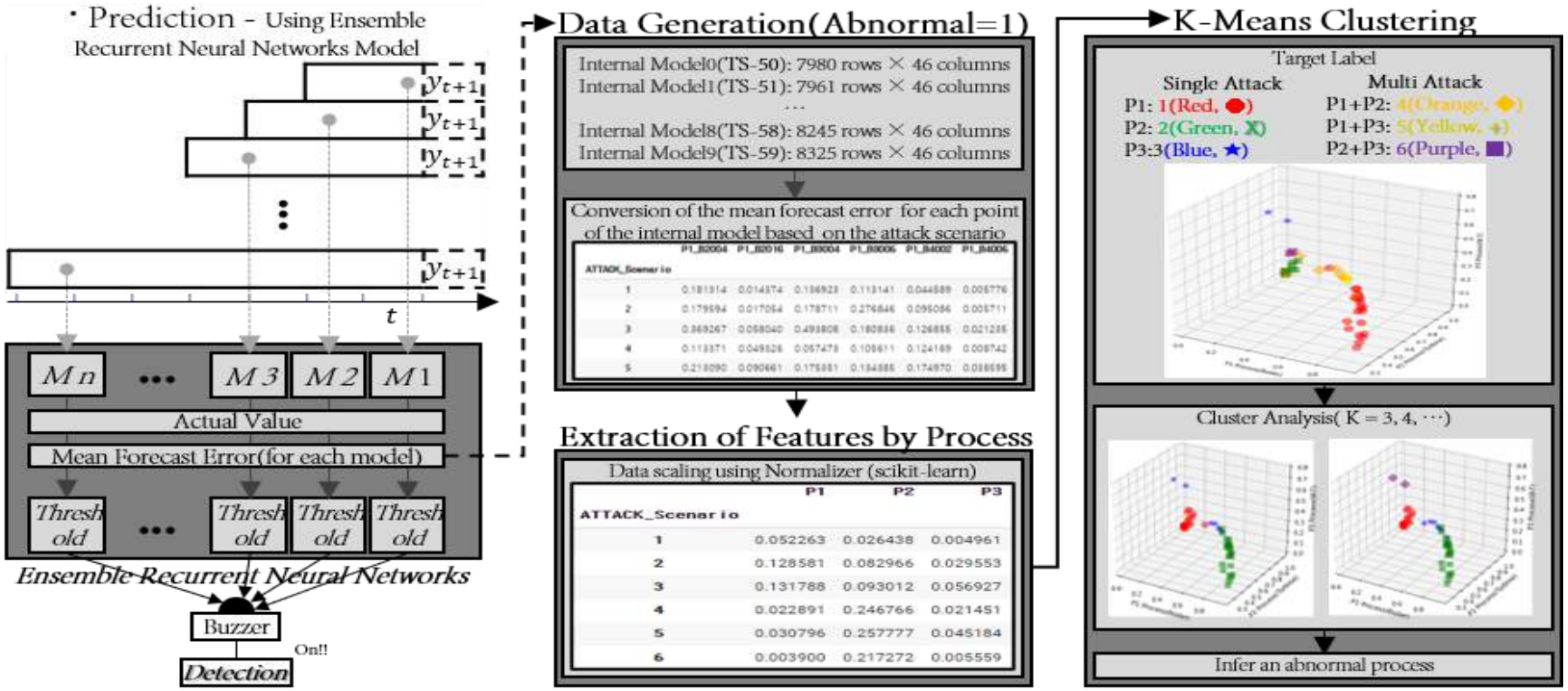
- 타임스텝 값의 조정(입력크기가 달라짐)과 학습데이터를 랜덤하게 하여 다수의 내부 모델을 생성함.

- 생성된 내부 모델들을 하나의 앙상블 모델로 표현하고 내부 모델 중 하나라도 비정상 탐지한다면, 현재의 공정이 비정상임을 나타냄.



4. 예측 오차를 통한 비정상 공정 도출

- 앙상블 모델을 구성하는 내부 모델들의 평균예측오차를 추출
- 공격 시나리오 기준으로 공정별 특징 추출(+데이터 전처리)
- K-평균 클러스터링을 이용하여 비정상 공정 도출



4-1. 예측 오차 데이터 생성

- 앙상블 모델을 구성하는 내부 모델들의 평균예측오차를 추출
- 공격 시나리오 기준으로 공정별 특징 추출(+ 데이터 전처리 (L2 Norm, *Normalizer()*)

ATTACK_Scenario	P1_B2004	P1_B2016	P1_B3004	P1_B3005	P1_B4002	P1_B4005	P1_B400B	P2_VT01	P2_VXT03	P2_VYT03	P3_FIT01	P3_LCP01D	P3_LCV01D	P3_LIT01	P3_PIT01
1	0.181314	0.014374	0.136923	0.113141	0.044589	0.005776	0.008552	0.093358	0.050935	0.088062	0.014054	0.015088	0.015361	0.018378	0.012305
2	0.179594	0.017054	0.178711	0.276846	0.095086	0.005711	0.010013	0.046807	0.035764	0.028124	0.018382	0.018587	0.020487	0.014751	0.023455
3	0.369267	0.058040	0.493808	0.180836	0.126855	0.021235	0.040594	0.108136	0.080433	0.082114	0.061747	0.063751	0.071088	0.056792	0.056103
4	0.113371	0.049326	0.057473	0.105611	0.124169	0.008742	0.015826	0.224880	0.098385	0.134702	0.039473	0.030922	0.059281	0.024136	0.049189
5	0.213090	0.090661	0.175351	0.134385	0.174970	0.038595	0.064461	0.455745	0.309935	0.210328	0.104782	0.154827	0.170785	0.089031	0.137136
6	0.654932	0.174796	0.302713	0.217384	0.306808	0.055369	0.115892	38.251709	1.599811	2.946987	0.275056	0.236618	0.258245	0.174578	0.182290
7	0.499504	0.144781	0.704358	0.389715	0.175348	0.044734	0.084165	0.358778	0.196522	0.221415	0.284600	0.293395	0.227157	0.171073	0.228820
45	0.679196	0.187215	0.238444	0.194122	0.264031	0.042543	0.119857	31.572697	1.486504	2.439092	0.525656	0.475315	0.517937	0.166855	0.263446
46	0.180618	0.031953	0.266935	0.177822	0.109886	0.009384	0.019743	0.763549	0.156193	0.158103	0.081442	0.050005	0.054584	0.031154	0.052002
47	0.676618	0.039418	0.059151	0.407713	0.113893	0.004706	0.006590	0.049395	0.023903	0.032556	0.011236	0.010274	0.016888	0.012129	0.009587
48	0.244380	0.071419	0.650316	0.207284	0.132003	0.025727	0.056189	0.121719	0.110030	0.087055	0.082197	0.097312	0.091032	0.067659	0.041666
49	0.703345	0.229879	0.271998	0.196001	0.274860	0.057348	0.122385	37.097295	0.611514	1.863674	0.434213	0.499639	0.422004	0.245772	0.459216
50	0.394192	0.052422	0.274889	0.251128	0.152265	0.015094	0.030063	0.334189	0.107207	0.218108	0.061835	0.061339	0.064299	0.072813	0.051756

ATTACK_Scenario	P1	P2	P3
1	0.052263	0.026438	0.004961
2	0.128581	0.082966	0.029553
3	0.131788	0.093012	0.056927
4	0.022891	0.246766	0.021451
5	0.030796	0.257777	0.045184
6	0.003900	0.217272	0.005559

$$X\_P1_1 = \frac{0.052263}{\sqrt{0.052263^2 + 0.026438^2 + 0.004961^2}} = \frac{0.052263}{0.058779} = 0.88914$$

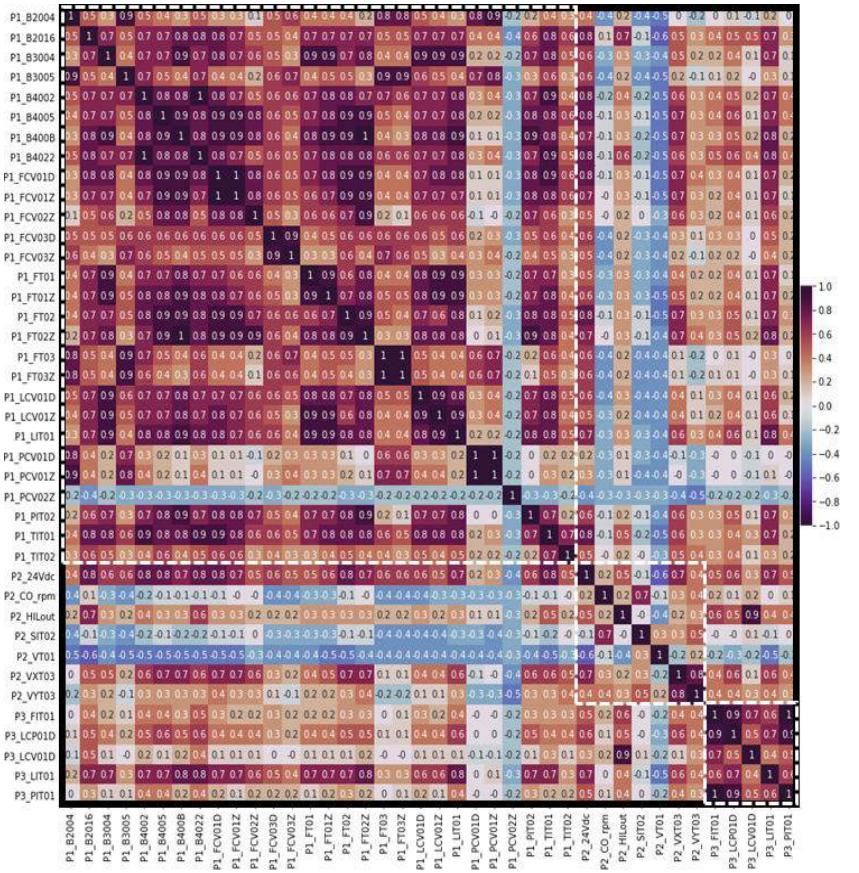
$$X\_P2_1 = \frac{0.026438}{\sqrt{0.052263^2 + 0.026438^2 + 0.004961^2}} = \frac{0.026438}{0.058779} = 0.44978$$

$$X\_P3_1 = \frac{0.004961}{\sqrt{0.052263^2 + 0.026438^2 + 0.004961^2}} = \frac{0.004961}{0.058779} = 0.08440$$

4-2. 예측 오차 데이터 분석

- 상관관계 분석: 동일 공정 내 Point 간 양의 상관관계가 확인, 일부 Point는 타 공정에서도 상관관계가 나타남

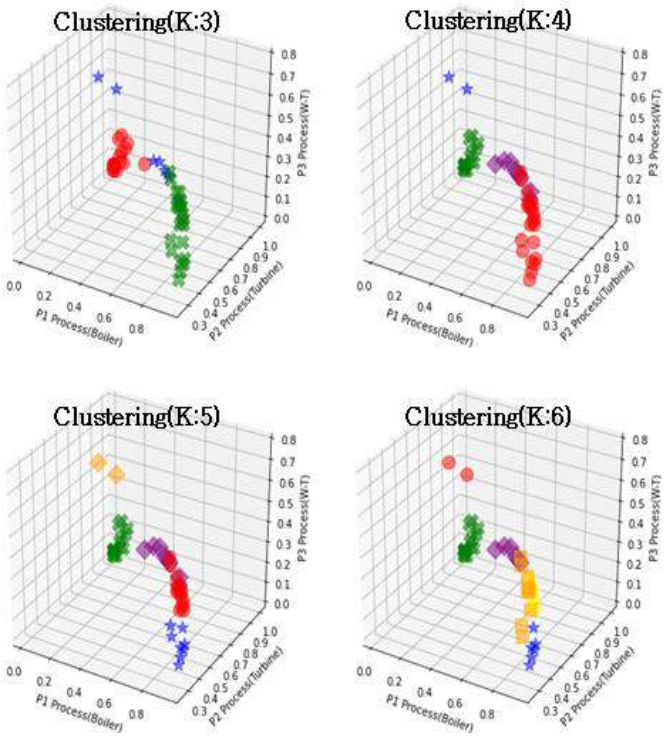
- 공격 시나리오별 공정의 평균 예측 오차 데이터를 생성하여 세부공정 (P1, P2, P3) 을 탐지하는 것이 적절함



Target Process	P1	P1	P1	P2	P2	P2	P1	P1	P1	P1	P1	P2	P1	P1	P1	P2	P2	P1	P1	P2	P1	P2	P1	P3	P3
Target Point	B2016	B6008	LCV01D	SCO	AutoSD	VTR02	B2016 PIT01	LCV01D	B6008 PIT08	B2016	PCV06D PIT06	VTR01	B6004 LIT01	B6004	PCV01D PIT01	AutoSD	AutoSD SIT01	LCV01D LIT01	PCV01D	SCO SIT01	PCV06D	RTR	PCV06D	LCP01D	LCV01D
Attack Point \	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25
P1_B2004	0.06	0.28	0.34	0.06	0.07	0.02	0.03	0.22	0.33		0.32	0.02	0.15	0.2	0.62	0.05	0.05	0.17	0.56	0.03	0.2	0.02	0.23	0.16	0.07
P1_B2016	0	0.03	0.05	0.03	0.03	0	0.01	0.05	0.05		0.02	0.01	0.04	0.06	0.04	0.02	0.03	0.06	0.04	0.03	0.03	0.01	0.03	0.03	0.06
P1_B3004	0.05	0.28	0.45	0.03	0.06	0.01	0.05	0.49	0.28		0.15	0.01	0.57	0.42	0.08	0.03	0.04	0.55	0.05	0.03	0.11	0.01	0.21	0.16	0.03
P1_B3005	0.04	0.48	0.17	0.06	0.05	0.01	0.03	0.25	0.33		0.14	0	0.13	0.2	0.33	0.04	0.02	0.13	0.36	0.03	0.29	0	0.16	0.18	0.06
P1_B4002	0.01	0.15	0.12	0.07	0.06	0.01	0.01	0.11	0.24		0.09	0	0.15	0.15	0.06	0.03	0.03	0.12	0.07	0.01	0.11	0.01	0.07	0.17	0.06
P1_B4005	0	0.01	0.02	0	0.01	0	0	0.01	0.02		0.01	0	0.01	0.02	0.01	0.01	0.01	0.01	0	0	0.01	0	0.01	0.02	0
P1_B400B	0	0.02	0.04	0.01	0.02	0	0.01	0.03	0.04		0.02	0	0.03	0.04	0.01	0.01	0.02	0.04	0	0.01	0.02	0	0.02	0.03	0.01
P1_B4022	0.01	0.08	0.08	0.03	0.03	0.01	0.01	0.07	0.14		0.04	0	0.08	0.08	0.03	0.01	0.02	0.08	0.04	0.02	0.05	0	0.04	0.09	0.07
P1_PCV01D	0	0.02	0.03	0.01	0.02	0	0.01	0.01	0.02		0.01	0	0.02	0.03	0	0.01	0.01	0.02	0	0.01	0.02	0	0.01	0.01	0
P1_PCV01Z	0	0.02	0.03	0.01	0.02	0	0.01	0.02	0.03		0.02	0	0.01	0.03	0	0.01	0.01	0.02	0	0.01	0.02	0	0.02	0.01	0
P1_PCV02Z	0	0.01	0.05	0	0.06	0	0.01	0.04	0.03		0.04	0.01	0.03	0.07	0	0.01	0.02	0.02	0	0	0.04	0.01	0.02	0.02	0
P1_PCV03D	0.02	0.31	0.3	0.02	0.02	0	0.03	0.32	0.27		0.65	0	0.24	0.48	0.14	0.02	0.01	0.24	0.15	0.01	0.7	0	0.51	0.37	0.02
P1_PCV03Z	0.02	0.26	0.2	0.04	0.04	0	0.02	0.14	0.31		0.56	0.01	0.1	0.25	0.17	0.02	0.02	0.11	0.21	0.02	0.42	0.01	0.35	0.27	0.02
P1_FT01	0.03	0.1	0.25	0.01	0.03	0	0.03	0.11	0.14		0.04	0	0.25	0.16	0.11	0.01	0.01	0.31	0.07	0.01	0.05	0	0.15	0.1	0.03
P1_FT01Z	0.01	0.09	0.12	0.02	0.03	0	0.02	0.06	0.16		0.02	0	0.13	0.13	0.05	0.01	0.01	0.16	0.05	0.01	0.04	0	0.07	0.09	0.01
P1_FT02	0	0.02	0.02	0.01	0.03	0	0.01	0.03	0.05		0.03	0	0.02	0.03	0.01	0.01	0.01	0.03	0	0.01	0.01	0	0.01	0.02	0
P1_FT02Z	0	0.01	0.03	0.01	0.03	0	0.01	0.03	0.03		0.03	0	0.03	0.04	0	0.01	0.02	0.03	0	0.01	0.02	0	0.02	0.02	0.01
P1_FT03	0.03	0.37	0.13	0.04	0.03	0	0.02	0.17	0.3		0.09	0	0.09	0.15	0.25	0.03	0.02	0.09	0.26	0.02	0.18	0	0.28	0.07	0.04
P1_FT03Z	0.02	0.39	0.11	0.03	0.03	0	0.02	0.14	0.18		0.07	0	0.09	0.14	0.23	0.03	0.02	0.07	0.24	0.02	0.14	0.01	0.25	0.1	0.03
P1_LCV01D	0.01	0.18	0.29	0.02	0.02	0	0.06	0.52	0.15		0.17	0	0.29	0.25	0.12	0.02	0.01	0.34	0.14	0.01	0.22	0	0.2	0.26	0.02
P1_LCV01Z	0.01	0.1	0.22	0.02	0.02	0	0.07	0.22	0.14		0.04	0	0.23	0.18	0.09	0.02	0.01	0.28	0.1	0.01	0.07	0	0.11	0.15	0.01
P1_LIT01	0.03	0.15	0.4	0.03	0.05	0	0.04	0.2	0.35		0.15	0	0.4	0.39	0.08	0.02	0.03	0.38	0.02	0.02	0.1	0	0.18	0.43	0.02
P1_PCV01D	0.05	0.1	0.06	0.01	0.01	0	0.02	0.09	0.08		0.03	0	0.08	0.07	0.37	0.01	0.01	0.09	0.39	0.01	0.05	0	0.04	0.1	0.03
P1_PCV01Z	0.04	0.15	0.06	0.03	0.02	0	0.02	0.12	0.1		0.04	0	0.08	0.09	0.36	0.02	0.01	0.09	0.38	0.02	0.09	0	0.05	0.08	0.02
P1_PCV02Z	0.99	0	0	0	0	0	0.98	0	0		0	0	0	0	0	0	0	0	0	0	0	0	0	0.01	0
P1_PIT02	0	0.01	0.03	0.01	0.02	0	0.01	0.03	0.05		0.04	0	0.02	0.05	0	0.03	0.03	0.04	0	0	0.01	0	0.02	0.04	0
P1_TIT01	0.01	0.05	0.06	0.03	0.02	0	0.01	0.05	0.07		0.03	0	0.05	0.05	0.01	0.01	0.01	0.05	0.01	0.01	0.04	0	0.03	0.03	0.03
P1_TIT02	0	0.01	0.02	0	0.02	0	0	0.02	0.01		0.01	0	0.02	0.01	0.01	0	0	0.01	0	0.01	0.01	0	0.02	0.02	0.01
P2_24Vdc	0.02	0.17	0.13	0.09	0.11	0.01	0.01	0.12	0.15		0.08	0.01	0.11	0.13	0.08	0.09	0.11	0.12	0.08	0.13	0.12	0.01	0.08	0.18	0.07
P2_CO_rpm	0.02	0.09	0.09	0.85	0.81	0.22	0.01	0.09	0.1		0.05	0.22	0.12	0.09	0.06	0.84	0.97	0.09	0.05	0.97	0.1	0.22	0.12	0.19	0.64
P2_HILout	0.02	0.09	0.12	0.08	0.06	0.01	0.01	0.08	0.11		0.06	0.01	0.08	0.11	0.07	0.06	0.06	0.09	0.06	0.07	0.09	0.01	0.07	0.09	0.45
P2_SIT02	0.04	0.07	0.07	0.46	0.49	0.23	0.02	0.06	0.09		0.03	0.23	0.1	0.07	0.04	0.48	0.08	0.06	0.03	0.09	0.06	0.23	0.12	0.09	0.2
P2_VT01	0.03	0.07	0.1	0.12	0.16	0.94	0.03	0.1	0.11		0.06	0.95	0.28	0.09	0.06	0.15	0.13	0.08	0.04	0.11	0.08	0.95	0.42	0.12	0.16
P2_VXT03	0.02	0.06	0.07	0.05	0.11	0.04	0.01	0.05	0.08		0.03	0.02	0.07	0.07	0.03	0.08	0.06	0.07	0.02	0.06	0.04	0.02	0.05	0.09	0.02
P2_VYT03	0.03	0.04	0.08	0.07	0.07	0.07	0.02	0.05	0.07		0.04	0.05	0.07	0.06	0.03	0.09	0.06	0.06	0.02	0.05	0.04	0.05	0.05	0.08	0.07
P8_FIT01	0	0.03	0.06	0.02	0.04	0.01	0.02	0.05	0.05		0.02	0.02	0.04	0.04	0.01	0.02	0.04	0.03	0.01	0.03	0.04	0.01	0.03	0.24	0.15
P8_LCP01D	0	0.03	0.06	0.02	0.05	0.01	0.02	0.05	0.05		0.02	0.02	0.05	0.04	0.01	0.02	0.03	0.05	0.01	0.03	0.03	0.01	0.03	0.2	0.09
P8_LCV01D	0.01	0.03	0.07	0.03	0.06	0.01	0.02	0.05	0.05		0.02	0.01	0.05	0.06	0.01	0.03	0.04	0.06	0.01	0.04	0.03	0.01	0.03	0.12	0.49
P8_LIT01	0.01	0.02	0.05	0.01	0.03	0	0.01	0.04	0.06		0.03	0	0.04	0.03	0.01	0.01	0.02	0.04	0.01	0.01	0.02	0	0.02	0.08	0.03
P8_PIT01	0	0.04	0.05	0.03	0.05	0	0.02	0.04	0.04		0.01	0.02	0.03	0.03	0.01	0.03	0.02	0.02	0.01	0.02	0.02	0.01	0.02	0.33	0.14

4-3. K-평균 클러스터링을 이용한 비정상 공정 도출

- K-평균 클러스터링을 이용하여 평균예측오차를 통해 비정상 공정을 도출함
- K값은 공정의 수를 최소값(3)으로 공격시나리오의 타입(단일:3, 복합:3 = 6)을 최대 값으로 사용(K = 3 ~ 6)



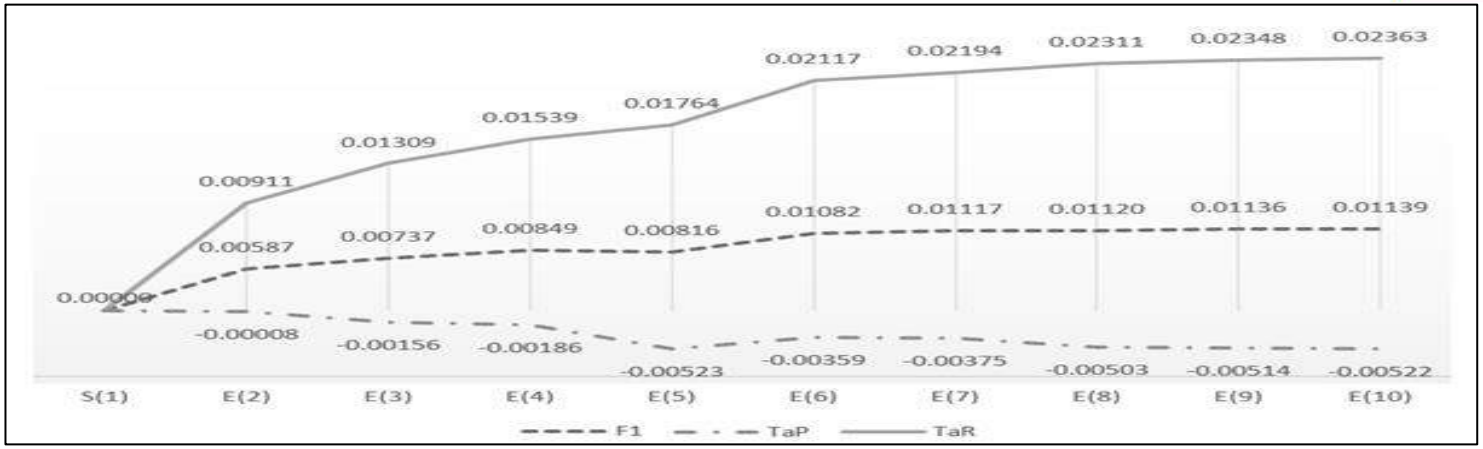
Target Process	Cluster		K:3		K:4		K:5		K:6	
	Result	Process	Data	Process	Data	Process	Data	Process	Data	Process
P1	1	P1	20	P1	19	P1	11	P1	11	P1
	2		2	P3	3	P1	3	P1	1	P1
	4						8	P1	8	P1
	5								2	P1
P2	0	P2	11	P2	11	P2	11	P2	11	P2
P3	2	P3	2	P3						
	3				2	P3	2	P3	2	P3
P1+P2	0	P2	4	P2	3	P2	3	P2	3	P2
	2	P1 P3	4	P3	5	P1	5	P1	5	P1
P1+P3	1	P1	3	P1	3	P1	3	P1	3	P1
P2+P3	0	P2	3	P2	3	P2	3	P2	3	P2

1. 기초, 내부, 앙상블 모델 탐지 성능 평가

- 딥러닝 알고리즘 성능 : 83.40% (RNN > CNN > AE)
- 기초모델(모델 튜닝) : 91.60%
- 내부모델(특징 선택) : 92.90%
- 앙상블 모델(내부모델 5개) : 93.80%
- 앙상블 모델(내부 모델 10개) : 94.04%

RNN(GRU) Model	Threshold	TaPR			
		F1-Score	TaP	TaR	Detected
Initial Model	0.075	**83.40%**	97.70%	72.70%	-
Basic Model	0.045	**91.60%**	97.90%	86.10%	47/50
Internal Model	0.045	**92.90%**	98.20%	88.30%	48/50
Ensemble Model (5 internal models)	0.045	**93.80%**	97.70%	90.20%	48(+1)/50
Ensemble Model (10 internal models)	0.045	**94.04%**	97.68%	90.66%	48(+1)/50

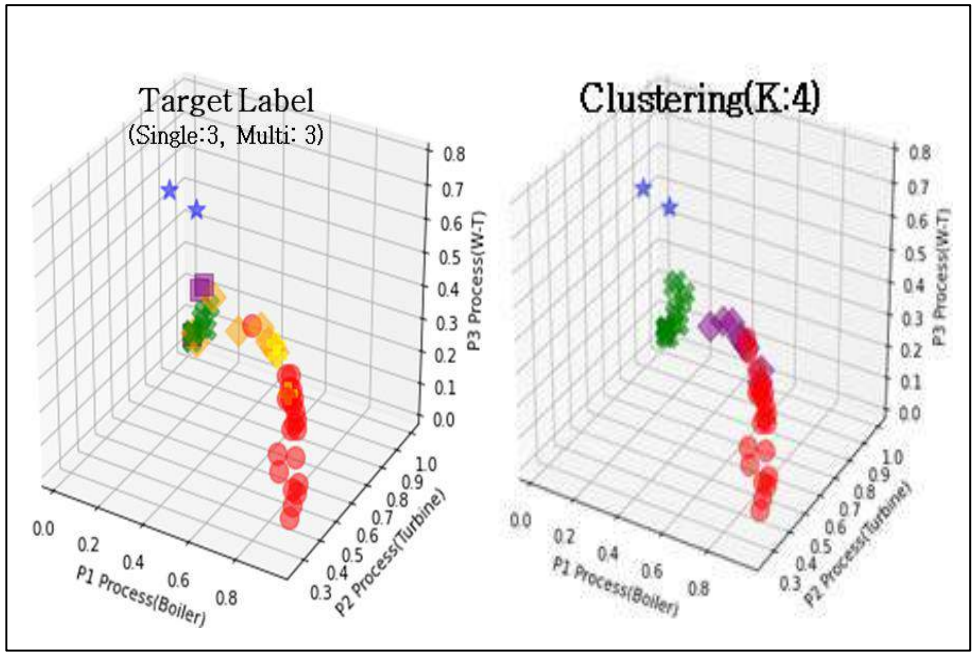
2. 내부 모델을 확장한 앙상블 모델 성능 비교



Ensemble / TimeStep		eTaPR(detect: 48)						CEM	
		F1		TaP		TaR		F1	
		Result	Inc/Dec	Result	Inc/Dec	Result	Inc/Dec	Result	Inc/Dec
S(1)	59	0.929	0	0.982	0	0.883	0	0.73645	0
E(2)	59 + 58	0.93487	0.00587	0.98192	-0.00008	0.89211	0.00911	0.74308	0.00663
E(3)	E(2) + 57	0.93637	0.00150	0.98044	-0.00148	0.89609	0.00398	0.74548	0.00240
E(4)	E(3) + 56	0.93749	0.00112	0.98014	-0.00030	0.89839	0.00230	0.74768	0.00220
E(5)	E(4) + 55	0.93716	-0.00033	0.97677	-0.00337	0.90064	0.00225	0.74864	0.00096
E(6)	E(5) + 54	0.93982	0.00266	0.97841	0.00164	0.90417	0.00353	0.75179	0.00315
E(7)	E(6) + 53	0.94017	0.00035	0.97825	-0.00016	0.90494	0.00077	0.75275	0.00096
E(8)	E(7) + 52	0.94020	0.00003	0.97697	-0.00128	0.90611	0.00117	0.75321	0.00046
E(9)	E(8) + 51	0.94036	0.00016	0.97686	-0.00011	0.90648	0.00037	0.75460	0.00139
E(10)	E(9) + 50	0.94039	0.00003	0.97678	-0.00008	**0.90663**	0.00015	0.75493	0.00033

3. 예측오차를 통한 비정상 공정 도출

- 단일공격 = P1: 1(Red, '●'), P2: 2(Green, 'X'), P3:3(Blue, '★'),
- 복합공격 = P1+P2: 4(Orange, '◆'), P1+P3: 5(Yellow, '+'), P2+P3: 6(Purple, '■')



Attack Scenario(50)		Clustering(K-means: 4)		
Type	Detected(49)	Data	Target	Result
Single	22	19	P1	1(P1)
		3		2(P1)
	11	11	P2	0(P2)
	2	2	P3	3(P3)
Multi	8	5	P1+P2	2(P1)
		3		0(P2)
	3	3	P1+P3	1(P1)
	3	3	P2+P3	0(P2)

4. 실험결과 분석

- 비정상 탐지
  - 딥러닝 모델 중 순환 신경망(GRU)을 이용한 모델이 적합
  - 앙상블 모델을 구성하는 오탐이 억제된 내부 모델은 공격이 끝난 지점 이후를 일정 기간(timestep) 비정상 탐지 하지만, 이 외에는 오탐지는 미 발생하였고 50개의 시나리오 중 49개를 탐지
  - 최종 성능은 94%이며, 기존에 발표된 논문의 성능보다 향상됨.
  - 미탐지에 대한 추가 탐지 방법이 강구되어야 함.

- 비정상 공정 도출
  - 앙상블 순환 신경망을 구성하는 내부 모델(단일)의 예측오차를 통한 비정상 공정을 도출하기 어려웠으며, 다수의 내부 모델의 오차 평균을 이용하는 것이 적절하게 클러스터링 되었음.
  - 단일 공격 및 복합 공격의 유형, 공격 횟수, 공격에 대한 파급력, 공정에 미치는 영향도 등을 달리하여 다양한 유형의 공격 시나리오가 추가된 상태에서 추가 연구가 필요해 보임.

1. 결론

- 비정상 공정을 탐지하여 운영 안정성을 보장하기 위한 연구를 진행하였음

  - 운전 데이터를 시계열 데이터로 변환하였으며, 회귀 예측 모델을 사용함.

  - GRU 기반으로 하는 비정상 탐지 모델은 오탐을 억제하여 정밀도(Precision)이 높은 상태를 유지함.

  - 버저 형태의 앙상블 순환 신경망을 구성하여 미탐지에 대해 상호보완적 동작하도록 함.

  - 비정상 탐지 후 예측 오차를 통한 비정상 세부 공정을 도출하여 대응 범위를 축소함.

2. 향후연구

- 데이터 측면
  - 다양한 정상 상황의 데이터를 수집해야 하며, 정상 상황에 대한 데이터 균형이 필요하다.
  - 정상 공정으로 복원 시, 시계열 데이터의 타임스텝 내에 비정상 데이터가 존재함으로 공격 종료 시점을 명확히 구분할 필요가 있다. (선택사항)
- 탐지 모델 측면
  - 오탐이 억제된 임계값을 자동으로 설정할 수 있는 구체적인 방안(자동화)이 필요하다.
  - 입력 데이터의 크기를 동적으로 받아들이고 내부적으로 학습할 수 있는 구조가 필요하다.
- 보안 측면
  - 비정상 탐지 후 비정상 공정에 대한 정의가 필요하다.(기계/전기 오류, 물리적 손상, 사이버 공격 등)
  - 비정상 탐지된 내용이 공정 운영 간 어느정도 심각도(상, 중, 하)를 갖는지 정의할 필요가 있다.

참고문헌

[CSO17] · and ." 을리 · · ··· 기술 연구 동향", ·· ·, 27(2), pp. 46-56, Apr. 2017.

[MIT19] MIT Technology Review, Triton is the world's most murderous malware, and it's spreading, 03, 2019. (last accessed 01-July-2021). [Online]. Available: https://www.technologyreview.com/2019/03/05/103328/cybersecurity-critical-infrastructure-triton-malware/

[KSX20] KS X IEC62443-4-2, " ·· · 보안 — 파 4-2부: ·· · ·보안 · ", 04, 2020.

[SHIN19] Hyeok-Ki Shin, Woomyo Lee, Jeong-Han Yun, and HyoungChun Kim, "Implementation of Programmable CPS Testbed for Anomaly Detection" 12th USENIX Workshop on Cyber Security Experimentation and Test (CSET19). Aug, 2019.

[SHIN20] Hyeok-Ki Shin, Woomyo Lee, Jeong-Han Yun and HyoungChun Kim, HIL-based Augmented ICS (HAI) Security Dataset, (last accessed 01-June-2020). [Online]. Available: https://github.com/icsdataset/HAI

[GOH17] Goh, Jonathan, et al. "Anomaly detection in cyber physical systems using recurrent neural networks." 2017 IEEE 18th International Symposium on High Assurance Systems Engineering (HASE). IEEE, 2017.

[DAN19] 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

[KEI15] Keith Stouffer, Victoria Pillitteri, Suzanne Lightman, Marshall Abrams and Adam Hahn, Guide to Industrial Control System Security, NIST SP 800-82, May. 2015.

[KMK17] ·, IoT, ICS에 대한 보안 · ▪ 및 · , · · 1813호 · · (IITP), 09, 2017

[TJW92] T.J.Williams, The Purdue Enterprise Reference Architecture, International Society of Automation, Research Triangle Park, North Carolina, ISBN 1-55617-265-6, 1992.

[LUC15] Luciana Obregon, Secure Architecture for Industrial Control Systems, SANS Institute, Sep, 2015.

[KSA18] ·. .·▪.· ··. ▪ 동향, · ··· · ·, pp. 31, 11, 2018.

[PAS19] · ·, ·· · 보안: · ICS 보안 가이드, (ISBN 9791161752389), 01, 2019.

[KEI20] Keith Stoufer, Victoria Pilliteri, Suzanne Lightman, Marshal Abrams and Adam Hahn(05. 2015), ·· · (ICS) 보안 가이드, (Translator), 05, 2020

[DON19] · and ." · · · · 기술 연구 동향", · ·, 56(8), pp. 3-12, Aug. 2019.

[CSN21] cs231n, CS231n Convolutional Neural Networks for Visual Recognition, cs231.stanford.edu. (last accessed 01-July-2021). [Online]. Available: https://cs231n.github.io/convolutional-networks/

[FEI17] Fei-Fei Li, Justin Johnson and Serena Yeung, Lecture 10: Recurrent Neural Networks, May. 2017. [Online]. Available: http://cs231n.stanford.edu/slides/2017/ cs231n_2017_lecture10.pdf

[SEPP97] Sepp Hochreiter and Jürgen Schmidhuber, "Long Short-Term Memory", Neural Computation, Vol. 9, No. 8, Nov, 1997

[CHR15] Christopher Olah, Understanding LSTM Networks, Aug, 2015. [Online]. Available: https://colah.github.io/posts/2015-08-Under standing-LSTMs/

[CHO14] 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 2014 Conference on Empirical Methods in Natural Language Processing, EMNLP2014, pp. 1724–1734, Sep, 2014.

[JON16] Jonathan Goh, Sridhar Adepu, Khurum Nazir Junejo and Aditya Mathur, "A dataset to support research in the design of secure water treatment systems" International Conference on Critical Information Infrastructures Security. 2016.

[ITR20] iTrust, Secure Water Treatment (SWaT.A7). May, 2020. [Online]. Available: https://itrust.sutd.edu.sg/itrust-labs_datasets/ dataset_info/

[KSM19] · ·and · , "SWaT · ·· · · 및 탐지 동향" ·· · ·, Vol. 29, No. 2, pp. 29-35, 4월. 2019.

[SHE15] Shengyi Pan, Thomas Morris and Uttam Adhikari, "Developing a hybrid intrusion detection system using data mining for power systems" IEEE Transactions on Smart Grid, Vol. 6, Issue. 6, pp. 3104-3113, Nov, 2015.

[UTT14] Uttam Adhikari, Shengyi Pan, and Tommy Morris, Industrial Control System (ICS) Cyber Attack Datasets. [Online]. Available: https://sites.google.com/a/uah.edu/tommy-morris-uah/ics-data-sets

[ANT16] Antoine Lemay and José M. Fernandez, Providing SCADA network data sets for intrusion detection research. Cyber Security Experimentation and Test (CSET 16), 2016.

[SND16] Antoine Lemay, SCADA network datasets. [Online]. Available: https://github.com/antoine-lemay/Modbus_dataset

[SHIN21] Hyeok-ki Shin, Woomyo Lee, Jeong-Han Yun, and Byung-Gil Min, "Two ICS Security Datasets and Anomaly Detection Contest on the HIL-based Augmented ICS Testbed", CSET'21: Workshop on Cyber Security Experimentation and Test, 2021

[HAI21] Hyeok-Ki Shin, Woomyo Lee, Jeong-Han Yun and HyoungChun Kim, HIL-based Augmented ICS(HAI) Security Dataset, (last accessed 01-July-2021). [Online]. Available: https://github.com/ icsdataset/hai/blob/master/hai_dataset_technical_details_v2.0.pdf

[CSO18] Seungoh Choi, Jeong-Han Yun, Sin-Kyu Kim, "A Comparison of ICS Datasets for Security Research Based on Attack Paths", In: Luiijf E., Žutautaitė I., Hämmerli B. (eds) Critical Information Infrastructures Security. CRITIS 2018. Lecture Notes in Computer Science, vol 11260. Springer, Cham.

[JUN17] Jun Inoue, Yoriyuki Yamagata, Yuqi Chen, Christopher M. Poskitt and Jun Sun, "Anomaly Detection for a Water Treatment System Using Unsupervised Machine Learning", 2017 IEEE International Conference on Data Mining Workshops (ICDMW), pp.1058-1065 Nov, 2017

[SIM18] Simon Duque Anton, Suneetha Kanoor, Daniel Fraunholz, and Hans Dieter Schotten, "Evaluation of Machine Learning-based Anomaly Detection Algorithms on an Industrial Modbus/TCP Data Set", Proceedings of the 13th International Conference on Availability, Reliability and Security, Aug, 2018

[ANT18] 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

[BER19] Giuseppe Bernieri, Mauro Conti, and Federico Turrin, "Evaluation of Machine Learning Algorithms for Anomaly Detection in Industrial Networks", 2019 IEEE International Symposium on Measurements & Networking (M&N) Measurements & Networking (M&N), pp. 1-6 Jul, 2019

[MOK21] 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

[KDY21] 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

[JON17] 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

[MOS18] 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, pp. 72-83, Jan, 2018

[KIM19] 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

[YIB19] Yibo Hu, Dinghua Zhang, Guoyan Cao and Quan Pan, "Network Data Analysis and Anomaly Detection Using CNN Technique for Industrial Control Systems Security", 2019 IEEE International Conference on Systems, Man and Cybernetics (SMC), pp. 593-597 Oct, 2019

[XING20] 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), Dec. 2020

[KYG21] Young Geun Kim, Jeong-Han Yun, Siho Han, Hyoung Chun Kim and Simon S. Woo, "Revitalizing Self-Organizing Map: Anomaly Detection using Forecasting Error Patterns," 36th International Conference on ICT Systems Security and Privacy Protection – IFIP SEC, pp. 382-397, Jun, 2021.

[NES18] Nesime Tatbul, Tae Jun Lee, Stan Zdonik, Mejbah Alam and Justin Gottschlich, "Precision and Recall for Time Series," 32nd Annual Conference on Neural Information Processing Systems (NeurIPS'18), Montreal, Canada, Dec, 2018.

[TSAD19] Nesime Tatbul, TSAD-Evaluator, Feb, 2019. (last accessed 01-July-2021). [Online]. Available: https://github.com/IntelLabs/ TSAD-Evaluator

[HWS19] 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 Addressing Ambiguous Labeling", CIKM'19: Proceedings of the 28th ACM International Conference on Information and Knowledge Management, pp. 2241-2244 Nov, 2019.

[TaPR19] Won-seok Hwang, TaPR, Nov, 2019. (last accessed 01-July-2021). [Online]. Available: https://github.com/saurf4ng/TaPR

[TaPR20] Won-seok Hwang, HAICon 2020 평가 지표 (TaPR) 설명, Jul, 2020. (last accessed 01-July-2021). [Online]. Available: https://www.slideshare.net/daconist/etapr-237428659

[KIM21] and ," ·· · . · 탐지," Journal of The Korea Institute of Information Security and Cryptology, 31(3), pp. 401-410, Jun. 2021.

[AAR16] Aaron van den Oord, Sander Dieleman, Heiga Zen, Karen Simonyan, Oriol Vinyals, Alex Graves, Nal Kalchbrenner, Andrew Senior and Koray Kavukcuoglu, "WaveNet: A Generative Model for Raw Audio," arXiv:1609.03499v2, Sep. 2016.

[COL21] "Colab Pro", google colaboratory, (last accessed 01-Mar-2021). [Online]. Available: https://colab.research.google.com/notebooks/pro.ipynb

Q & A

감사합니다.