

2022 KNS 추계 – 워크숍 M
원자력발전소 운전지원시스템 개발

Strategy to coordinate actions for multi-agent reinforcement learning environment



Oct 19th, 2022

김재민

Ulsan National Institute of Science and Technology

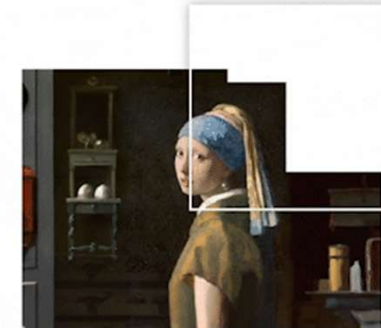
Address 50 UNIST-gil, Ulju-gun, Ulsan, 44919, Korea

Tel. +82 52 217 2371 **Web.** psa.unist.ac.kr



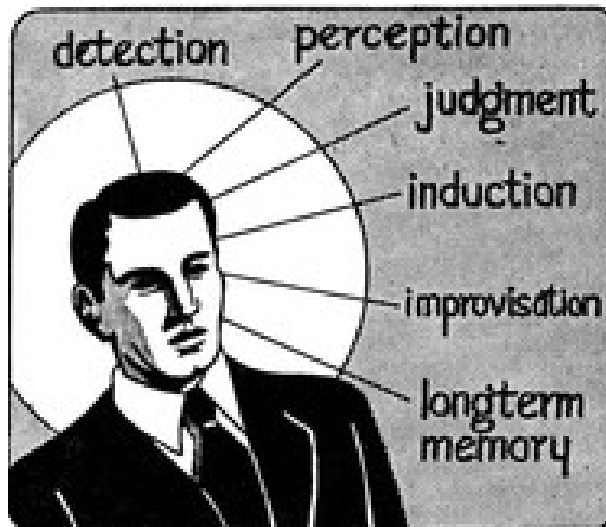
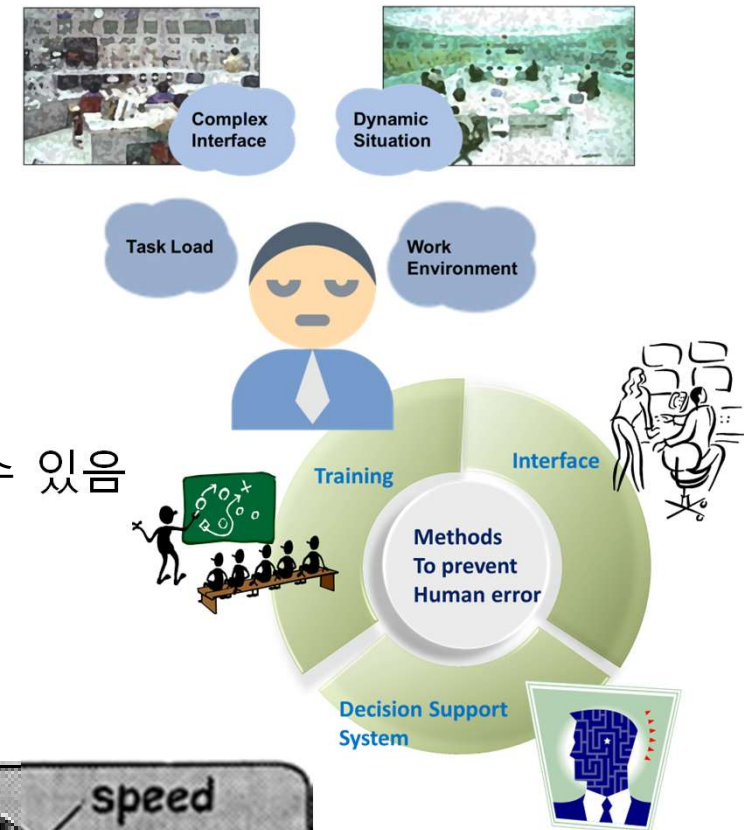
NUCLEAR Safety Assessment and Plant HMI Evolution

인공지능 기반 지원시스템

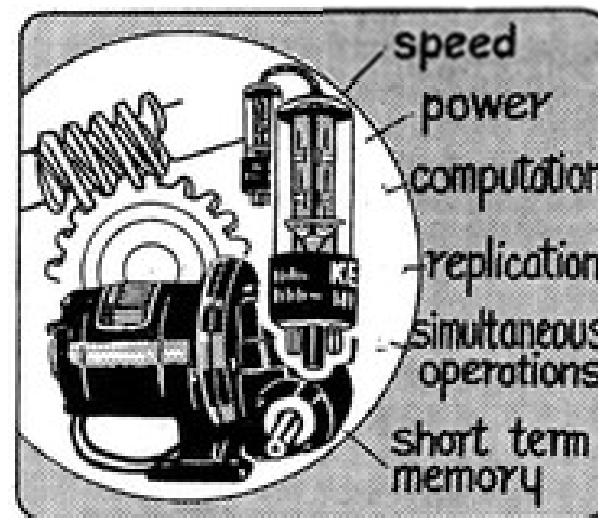


원전 운전지원시스템

- 원전 산업은 안전에 대한 높은 기준이 요구됨
- 운전원은 수많은 정보를 다루기 위해 고도의 훈련을 받음
- 새로운 시스템을 신뢰하고 적용하기 위한 검증이 필요함
- 인간의 대체가 아닌 지원 도구로 인공지능 기술을 활용할 수 있음



Human At Better At



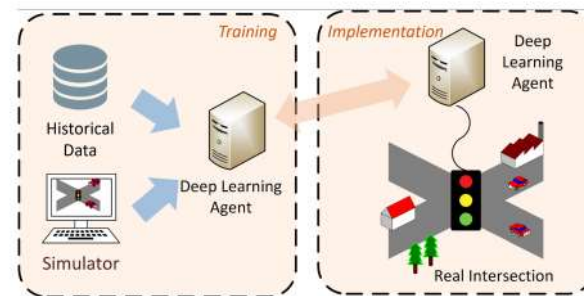
Machine At Better At

원전 운전지원시스템

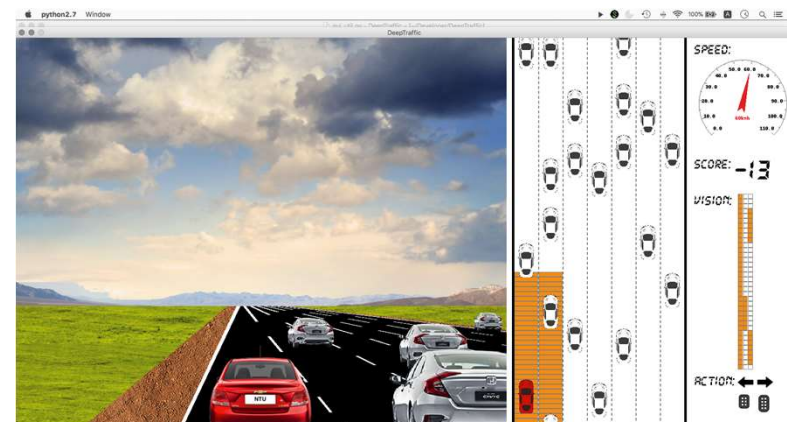
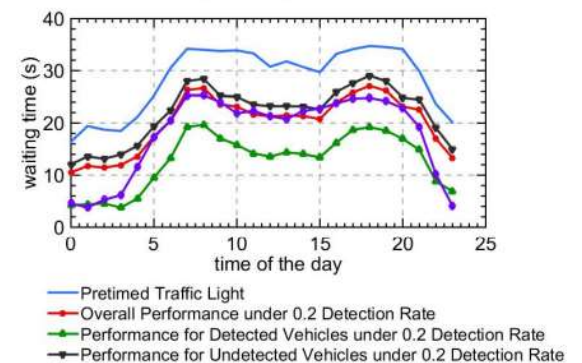
- 소형모듈원전과 같은 차세대 원전은 운전 및 운영 효율을 위해 운전원 그룹의 수를 줄임
 - 3명의 운전원이 여러 대의 원자로를 운영함
- 운전원 업무부하 저감 등을 위해 자동운전 시스템, 운전지원시스템 등이 도입될 것임
- 강화학습을 통해 정형화하기 힘든 정책을 습득하여 자동운전을 구현하고자 함
 - Traffic control, autonomous vehicle



KAERI SMART 주제어실 예시

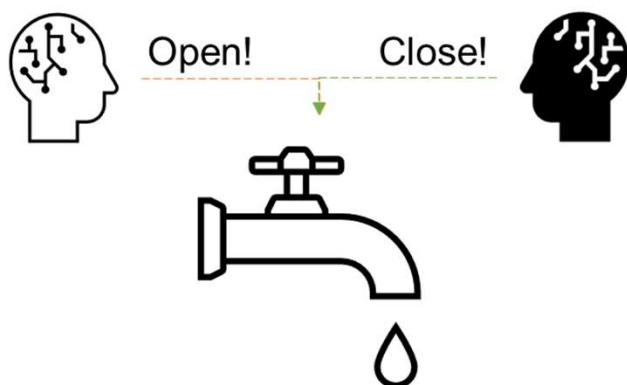


Performance of proposed system at different time of a day



인공지능 적용시 고려할 사항

- 원전과 같은 큰 규모의 시설은 수많은 기기와 계통들이 구성되어 있음
- 강화학습 환경 구성시 한번에 모두 고려하면 차원이 너무 커져서 학습 효율이 떨어짐
- 목적에 맞는 에이전트를 따로 학습시켜 Multi-agent 환경으로 구성하는 것이 유리함
- 에이전트간 행동이 충돌하여 반대되는 행동을 수행하려고 할 때 이를 해결할 방법이 필요함
- 현재 원전에는 어떤 행동이 우선되어야 하거나 다른 새로운 행동이 필요하다는 근거로 쓰일 정량적 지표가 없음
- 이를 해결하기 위하여 에이전트간 행동을 조율하기 위한 방법론을 제시함

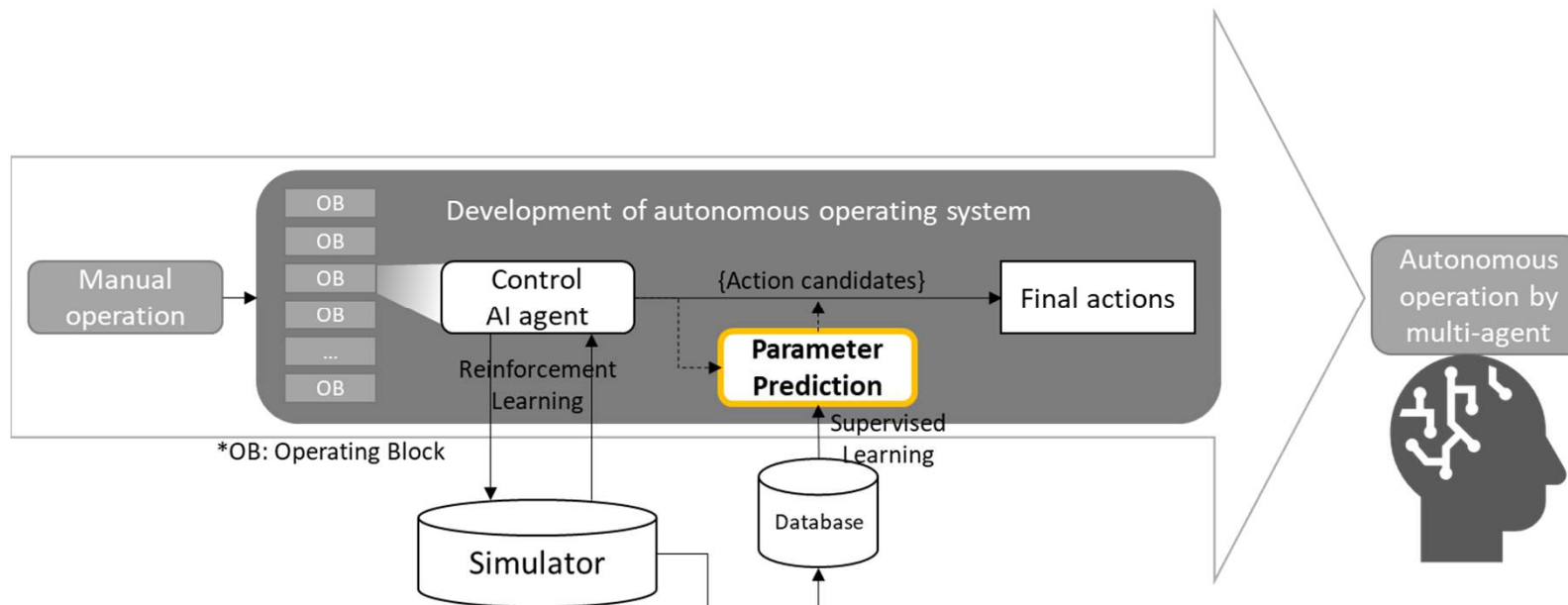


≠



기동운전 자동화

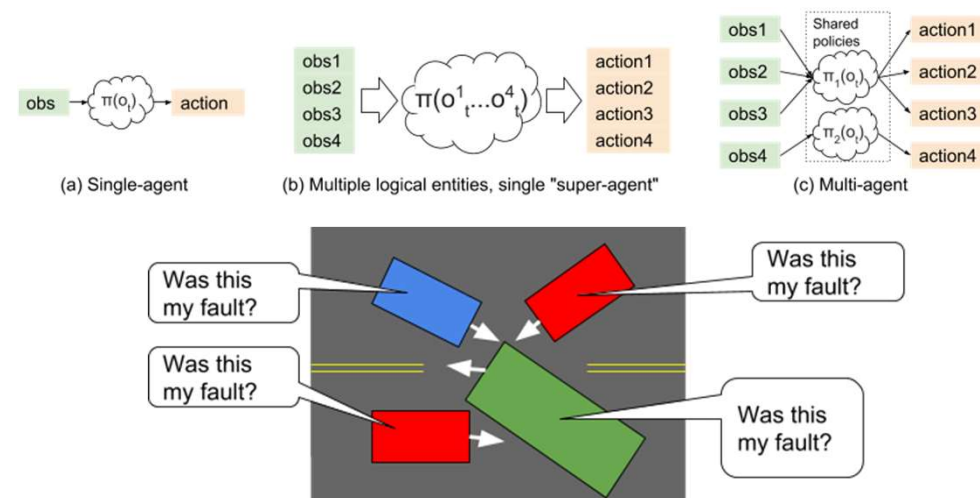
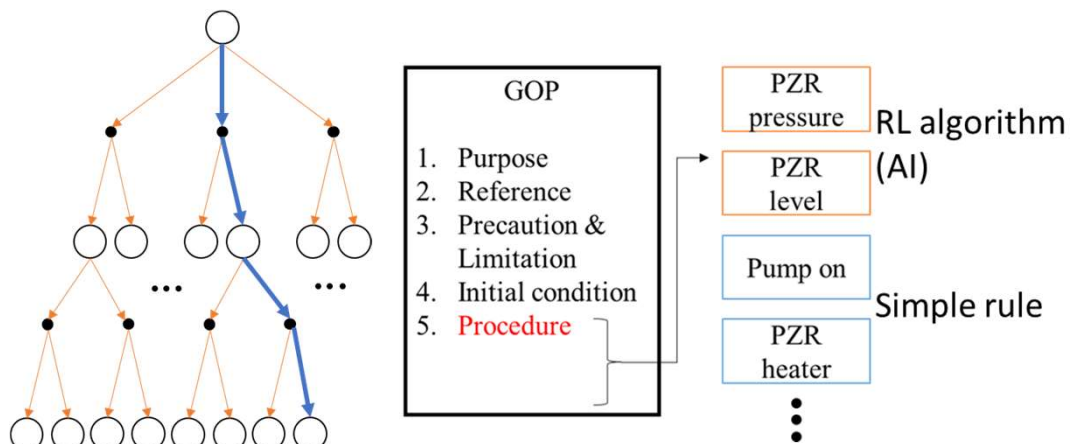
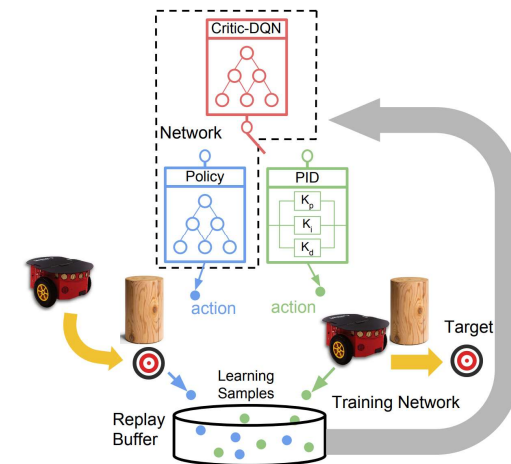
- 운전 직무를 세분화하여 인공지능 에이전트가 다루는 정보의 양을 줄임
- 여러 에이전트가 동시에 수행될 때 행동 충돌을 고려하여 변수 예측 모델을 사용함
- 행동 충돌 상황에서 최종 행동 선정을 위해 운전제한조건(LCO)을 기준으로 사용함
- 재학습 없이 에이전트간 행동을 조율하여 최적 행동을 찾을 수 있도록 함



인공지능 모델

❖ 강화학습

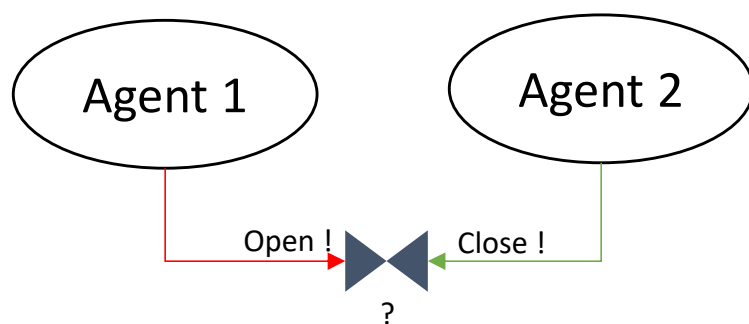
- Soft actor critic (SAC)
- 강화학습을 통해 습득한 운전정책을 가진 인공지능 에이전트
- 강화학습시 인공지능 에이전트가 모든 에피소드를 체험하는 것은 비효율적임
 - 정책을 수렴시키기 위한 탐험과 착취를 적절히 수행함 (exploration & exploitation)
 - 하나의 에이전트로 모든 상태와 행동을 다루면 차원이 커져서 충분한 학습을 거쳤다고 하기 힘들
- 효율적인 학습을 위해 필요한 정보를 가진 운전블록으로 구성



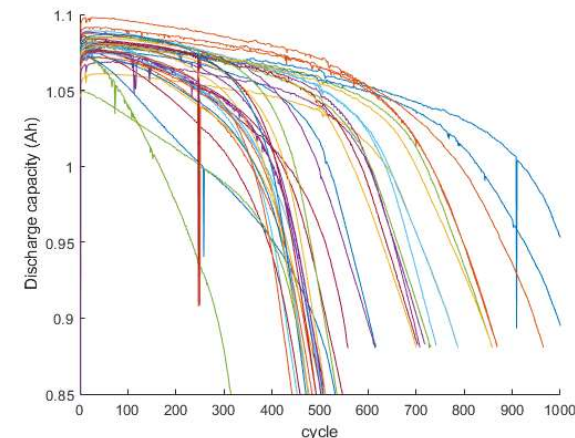
인공지능 모델

❖ 변수예측모델

- Long short-term memory (LSTM)
- 회귀 모델로 변수의 미래 상태를 예측하는 역할을 수행함
- 선택된 행동의 가치를 평가하기 위한 예측값을 제공함



close open close open
0000
 FV122 HV142

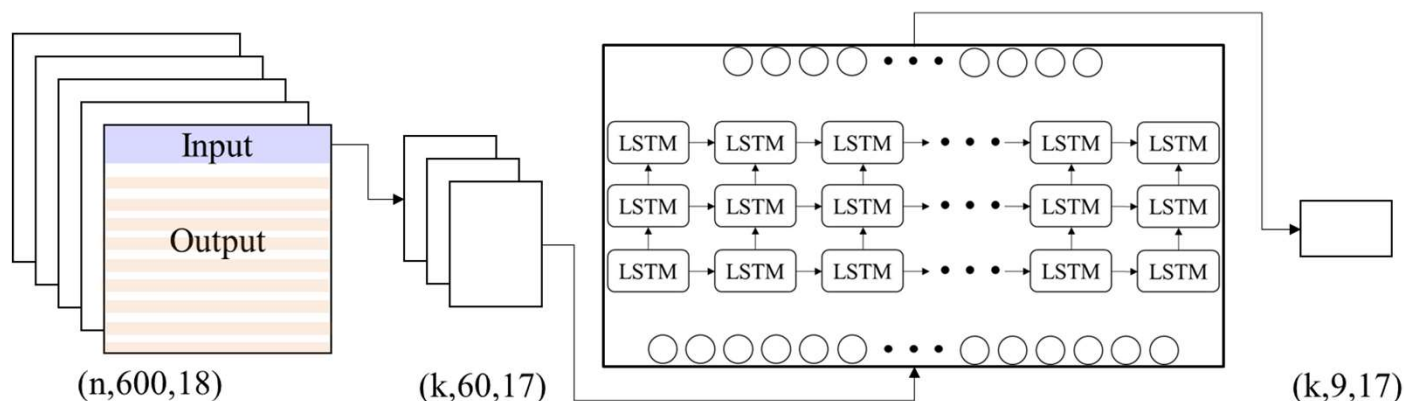


| | | | | | |
|------|-------------|------|--------------------------|------|---------------------------|
| 0000 | No action | 0100 | FV122 open | 1000 | FV122 close |
| 0001 | HV142 open | 0101 | FV122 open & HV142 open | 1001 | FV122 close & HV142 open |
| 0010 | HV142 close | 0110 | FV122 open & HV142 close | 1010 | FV122 close & HV142 close |

인공지능 모델

❖ 변수예측모델

- 예측모델은 17개 변수로 학습됨 (LCO 기반 변수 선정 + RCS 온도)
- 10분 간격으로 랜덤 행동을 주입한 100분의 시나리오 구성함
 - 총 1000개의 시나리오를 구성하여 10,000개의 데이터셋을 사용함
- 회귀모델은 처음 1분까지의 데이터를 입력으로 받고 이후 9분간 변수 상태를 예측
- 예측모델 학습결과 지표: RMSE

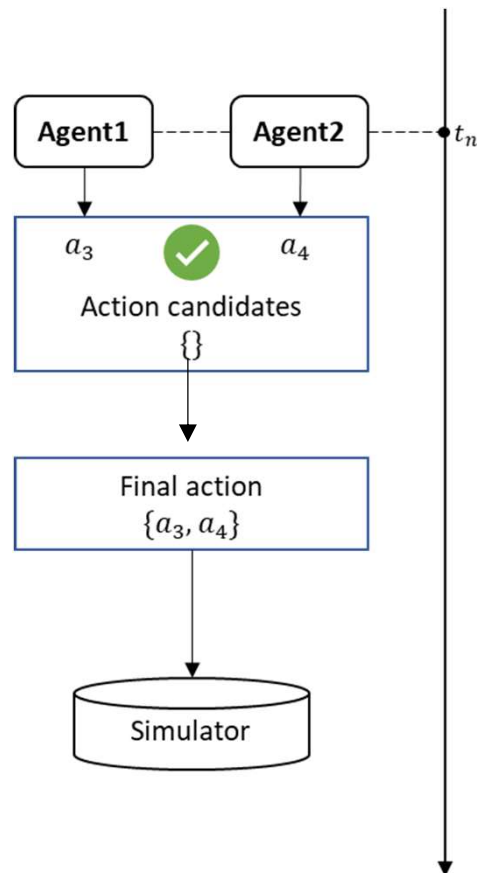


| No. | Description |
|-----|------------------------------|
| 1 | PZR TEMPERATURE. |
| 2 | LETDOWN BACK PRESSURE |
| 3 | LETDOWN FLOW |
| 4 | CHARGING FLOW |
| 5 | PZR PRESSURE(NARROWRANGE) |
| 6 | LOOP 3 AVERAGE TEMP |
| 7 | LOOP 2 AVERAGE TEMP |
| 8 | LOOP 1 AVERAGE TEMP |
| 9 | PZR LEVEL |
| 10 | VOLUME CONTROL TANK LEVEL. |
| 11 | VCT PRESSURE |
| 12 | RCP SEAL INJECTION FLOW |
| 13 | RCP SEAL NO.1 DELTA PRESSURE |
| 14 | RCP SEAL NO.1 RETURN FLOW |
| 15 | S/G 3 PRESSURE |
| 16 | S/G 2 PRESSURE |
| 17 | S/G 1 PRESSURE |

행동 후보와 변수 예측 모델

❖ Case 1: 행동충돌이 발생하지 않은 경우

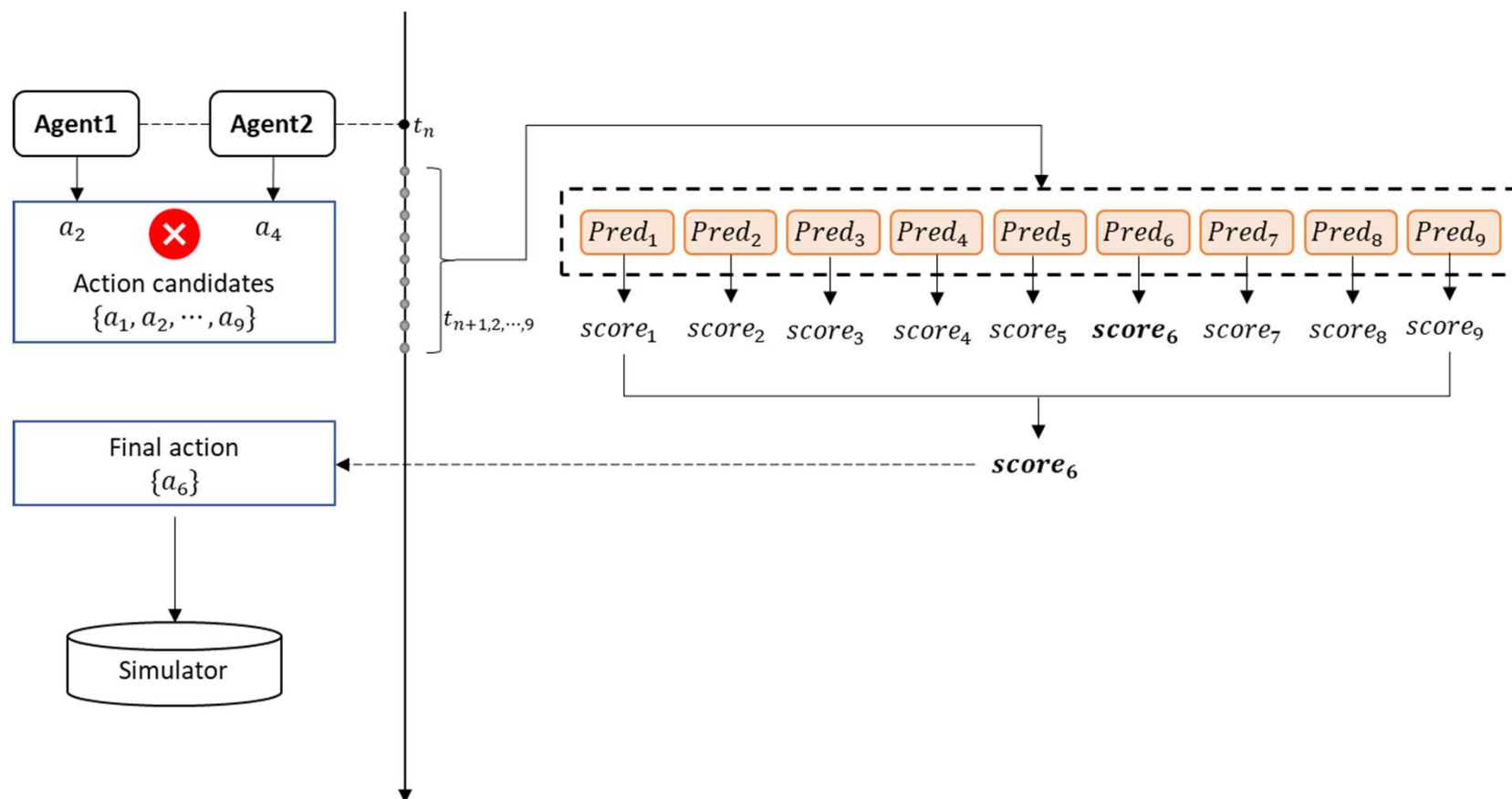
- 운전블록의 행동이 그대로 수행됨



행동 후보와 변수 예측 모델

❖ Case 2: 행동충돌이 발생한 경우

- 상충되는 행동이 발생한 순간까지 측정된 값을 기준으로 예측 모델을 활성화시킴
- 예측된 값을 정량화하여 가장 높은 점수를 얻은 모델에 할당된 행동을 선택함
- 행동 후보는 에이전트의 행동들만 고려하거나 가능한 모든 행동의 조합을 고려할 수 있음



Quantification strategy

1. Comparison: 단순 비교

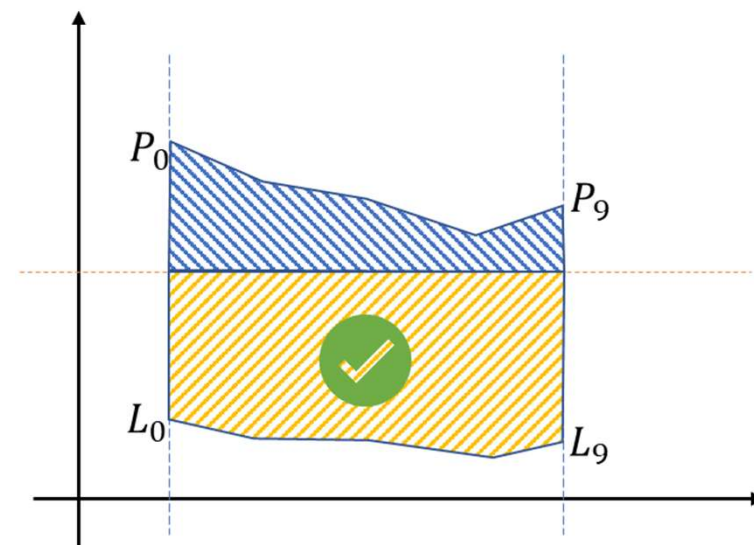
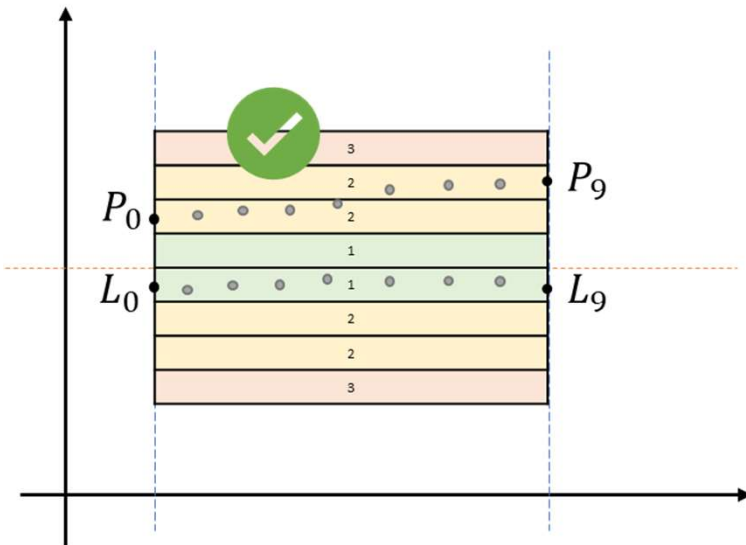
$$Score_{comp}: P_9 > L_9$$

2. Average: 예측 순간들의 평균값을 비교

$$Score_{avg}: P = \frac{P_1 + P_2 + P_3 + \dots + P_9}{9}, L = \frac{L_1 + L_2 + L_3 + \dots + L_9}{9}$$

3. Area: 예측값과 목표값 사이의 면적을 비교

$$Score_{area}: P = \sum \frac{(P_i + P_{i+1})}{2}, L = \sum \frac{(L_i + L_{i+1})}{2}$$

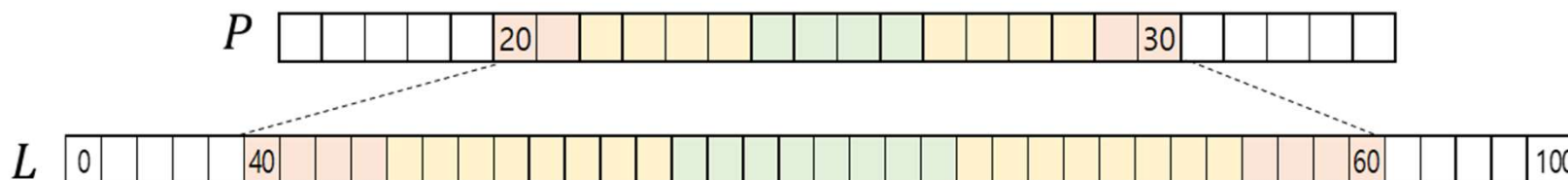
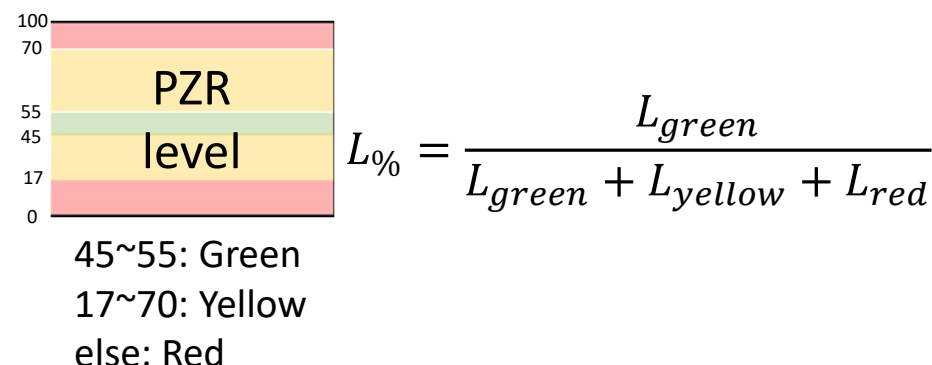
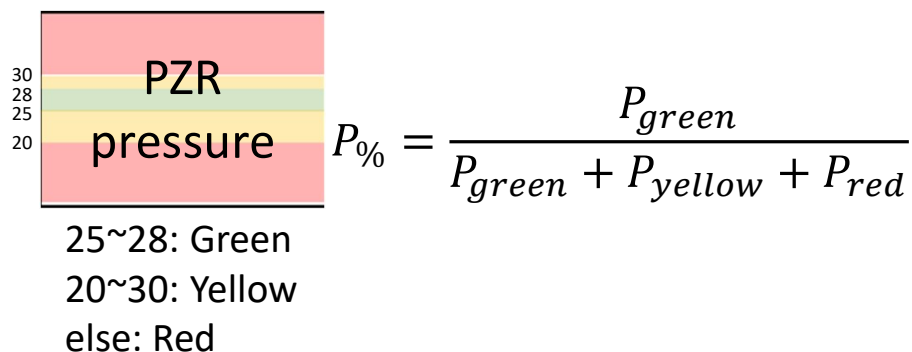


Quantification strategy

- 최종 점수는 각 지표에 대한 점수의 합으로 나타냄
- 가열율과 같이 위반시 시나리오를 중지해야할 지표는 -10 또는 0의 값을 주도록 설정함

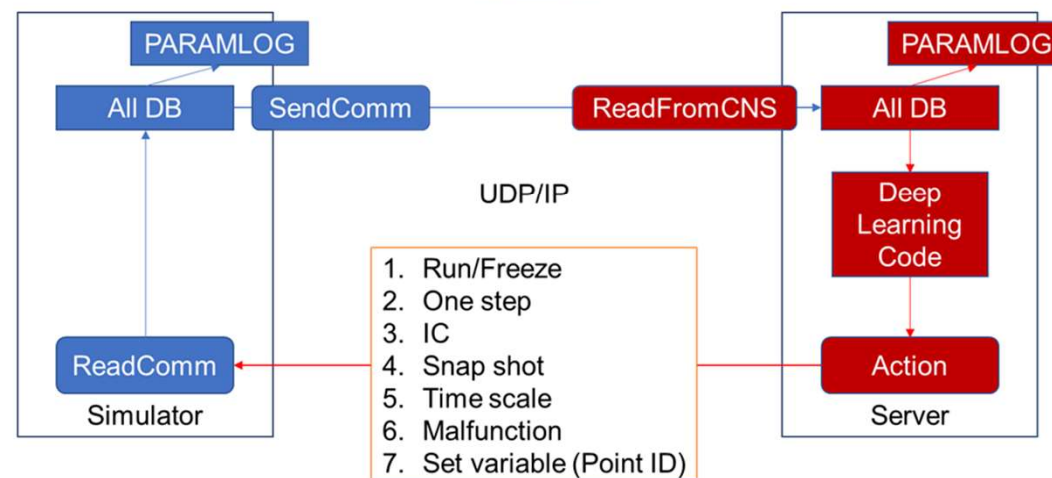
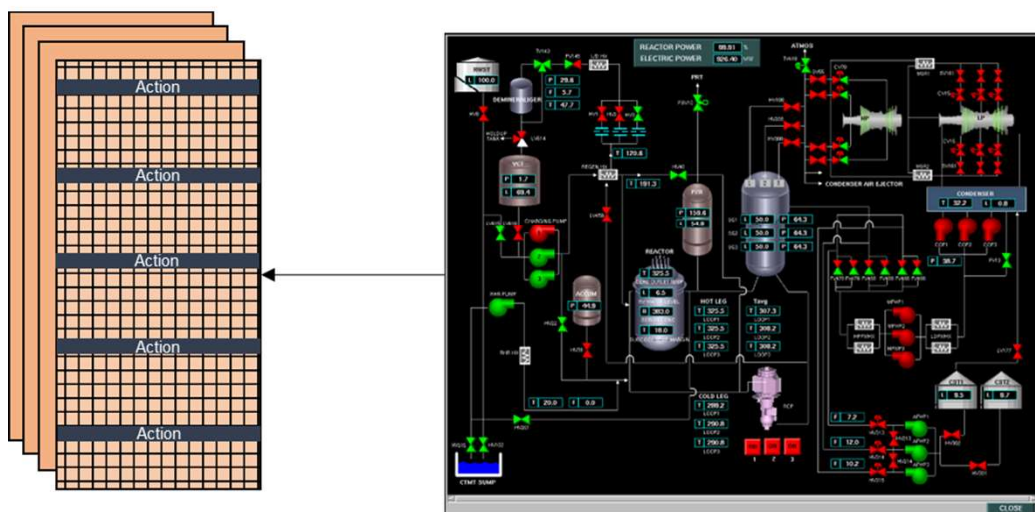
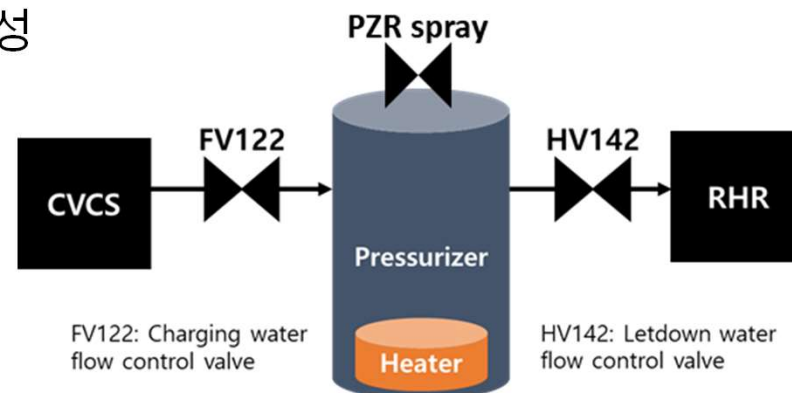
$$score_{total} = score_{heatrate} + score_{pressure,k} + score_{level,k}$$

- 운전수행척도는 LCO를 기반으로 변수 범위를 나누어 운전 중 녹색 범위에 오래 머문 비율을 기준으로 정함
- 점수 계산 시 scale을 고려하여 조정함



학습환경 구성

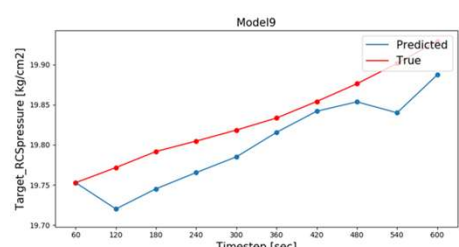
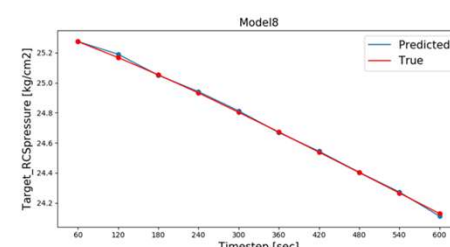
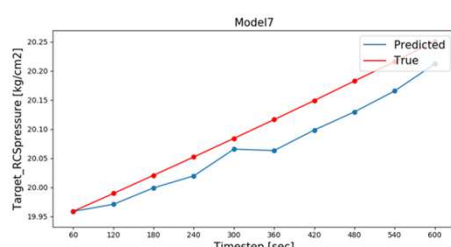
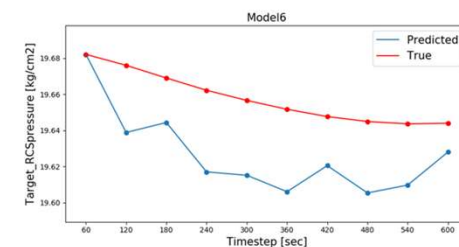
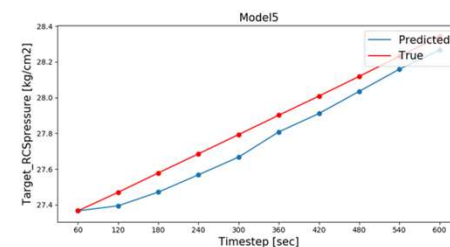
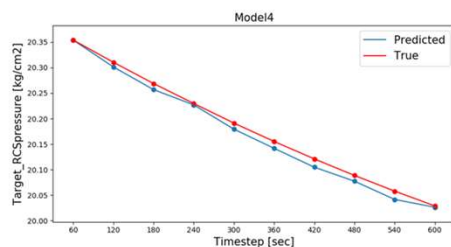
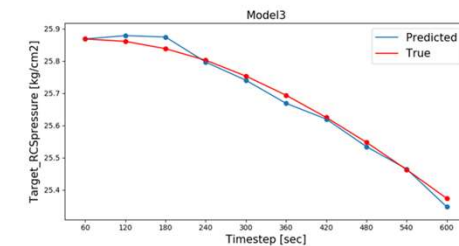
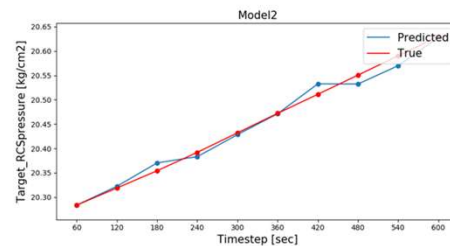
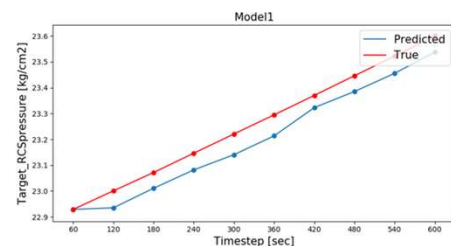
- Compact Nuclear Simulator 사용 (993MWe WH 3-loop PWR)
- 상온정지부터 고온대기까지 운전구간
- 행동충돌은 가압기 기포형성 이후 가압기의 압력과 수위를 조절할 때 발생 가능
- 강화학습은 시뮬레이터와 연동된 학습 환경을 구성
- 예측모델 훈련데이터는 시나리오대로 생산



결과

- 9개 행동에 대한 예측 모델의 학습 결과 (표)
- 특정 순간에 수행할 행동에 따른 변수 예측 결과 (그래프)
- Scale을 고려할 때 완전 일치하지 않아도 경향 파악 가능

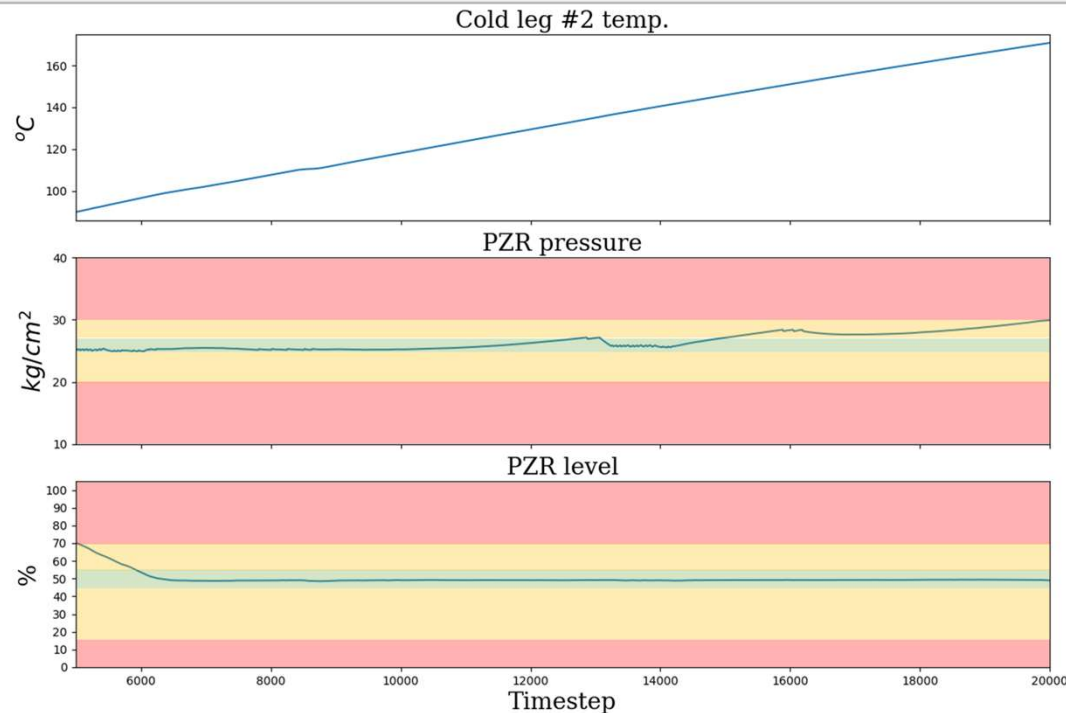
| Model (action) | Train_score RMSE | Val_score RMSE |
|----------------|------------------|----------------|
| 0000 | 0.16440 | 0.16498 |
| 0001 | 0.17851 | 0.18993 |
| 0010 | 0.16424 | 0.16695 |
| 0100 | 0.28892 | 0.28315 |
| 0101 | 0.12891 | 0.12905 |
| 0110 | 0.12082 | 0.12033 |
| 1000 | 0.14195 | 0.14212 |
| 1001 | 0.17615 | 0.17994 |
| 1010 | 0.11743 | 0.11613 |



결과

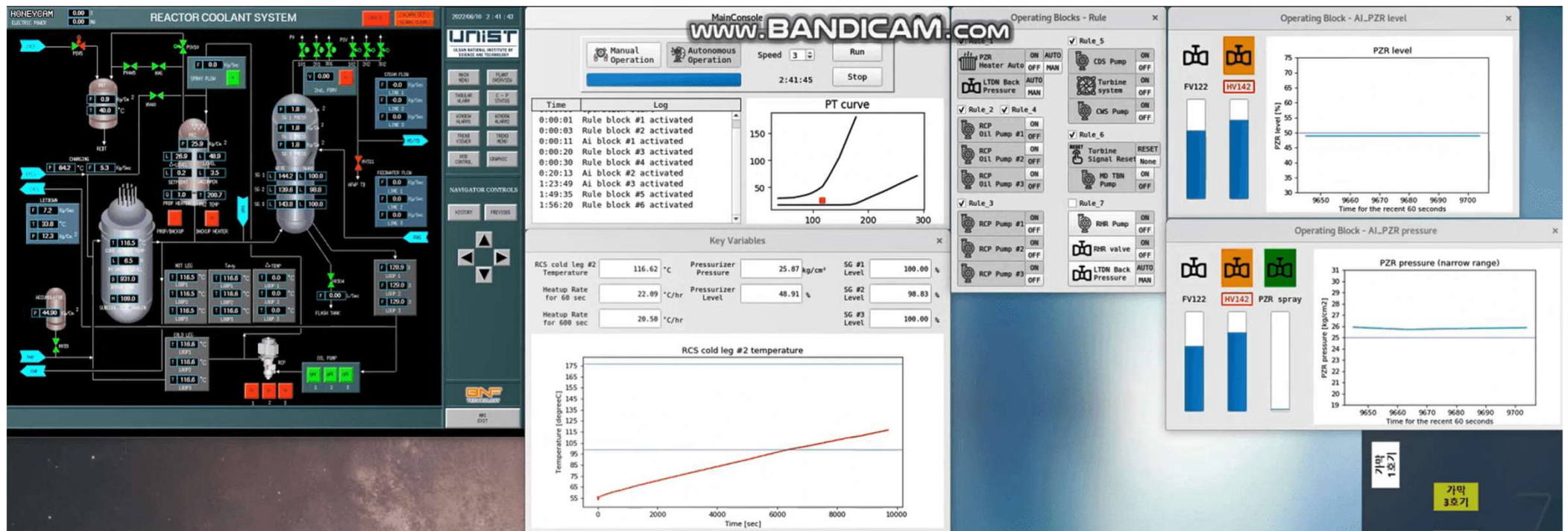
- 가압기 수위 70%부터 고온대기상태까지 운전한 결과

| Method | PZR pressure | | | | PZR level | | | |
|---------------|--------------|--------------|-----------|----------|-------------|--------------|-----------|----------|
| | P_{green} | P_{yellow} | P_{red} | $P_{\%}$ | L_{green} | L_{yellow} | L_{red} | $L_{\%}$ |
| 0. Raw | 9974 | 5257 | 0 | 65.5% | 14344 | 887 | 0 | 94.2% |
| 1. Comparison | 11790 | 3251 | 176 | 77.5% | 8640 | 6576 | 0 | 56.8% |
| 2. Average | 12039 | 2974 | 0 | 80.2% | 14084 | 929 | 0 | 93.8% |
| 3. Area | 12173 | 3085 | 0 | 79.8% | 14004 | 1254 | 0 | 91.8% |



데모영상

- 기동운전 중 인공지능 모델간 행동조율 영상
- 행동 충돌시 노란색으로 표시 > 최종 선택된 행동을 아래 박스로 표시
- 가압기 압력 및 수위를 적절한 수준에서 유지 가능



A satellite view of the Earth at night, showing the continents of Africa, Europe, and Asia. The landmasses are dark, while the oceans are a deep blue. Numerous bright white lights from cities and towns are visible across the land, particularly concentrated in Europe and Asia. The background is a dark, starry space.

THANK YOU

