

강화학습을 활용한 i-SMR 부하추종 운전 최적화 알고리즘 개발

이용준(원자력공학과), 장현종(인공지능학과)

jkh7542@khu.ac.kr

lezelamu@naver.com

경희대학교 일반대학원 총학생회
KYUNG HEE UNIVERSITY GRADUATE STUDENTS' ASSOCIATION

서 론

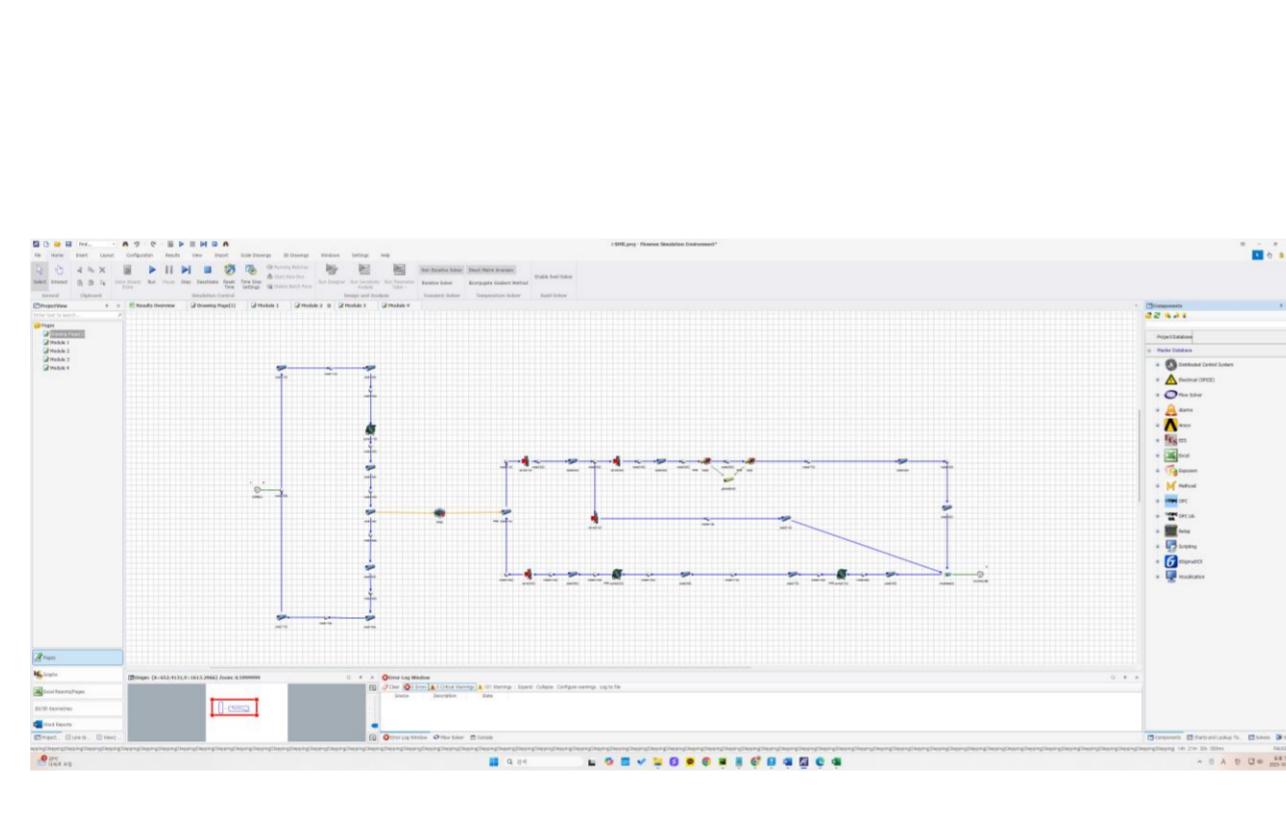
- 원자력발전은 전력 수요를 안정적으로 공급하는 기저부하전원으로서 핵심적인 역할을 지속해서 수행해 왔음
- 그러나 최근 재생에너지 보급 확대와 전력시장 구조변화로 인해 원자력발전에도 보다 유연한 출력 조절 능력이 요구되고 있음
- 부하추종운전은 전력계통의 안전성과 경제성을 동시에 확보하기 위한 주요 운전모드로 부각되고 있음 하지만 한국은 원전이 오랫동안 기저부하 정원의 역할에 집중해 왔으며, 경제성과 안전성 측면에서 탄력운전을 통한 상시 출력변화보다는 정출력 원전에 집중해 왔기 때문임
- 이러한 상황에서 한국에서도 차세대 원자로로써 i-SMR이 개발되어지고 있으며 SMR은 안전성, 경제성 그리고 유연성 측면에서 강점을 자기는 원자로이며 i-SMR에서 강조되는 10대 기술 중 **자동화와 재생에너지 연계**에 초점을 맞추어 강화학습을 활용한 i-SMR 부하추종 운전 최적화 알고리즘을 개발이 본 연구의 목표임

연구 내용

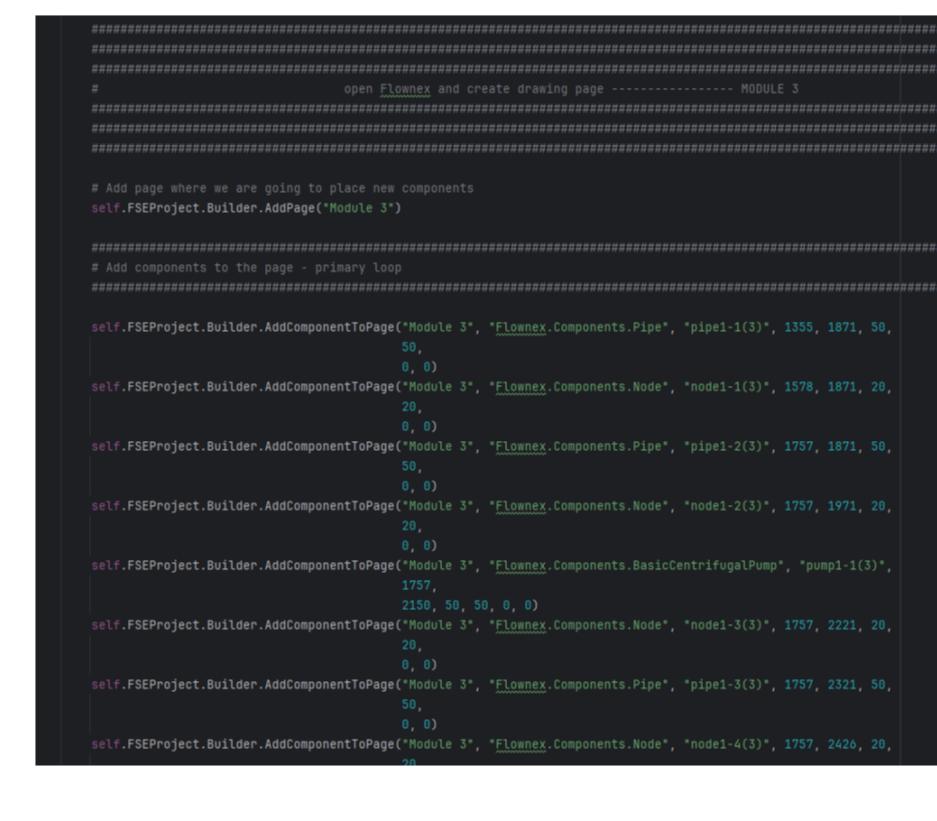
- Flownex SE는 열유체 시스템의 설계와 해석을 위한 포괄적인 시뮬레이션 환경을 제공하는 소프트웨어로 다수모듈 간 상호작용을 모사할 수 있고, Application Programming Interface를 지원하기에 본 연구에 활용
- 공개된 문헌을 기반으로 1차측과 2차측의 기본구성을 단순화하여 반여하였고, 170MWe 수준에 맞추기 위해 주요온도, 압력, 유량, 출력조건 등을 고려하여 Python API를 활용한 SMR 모델링을 수행함



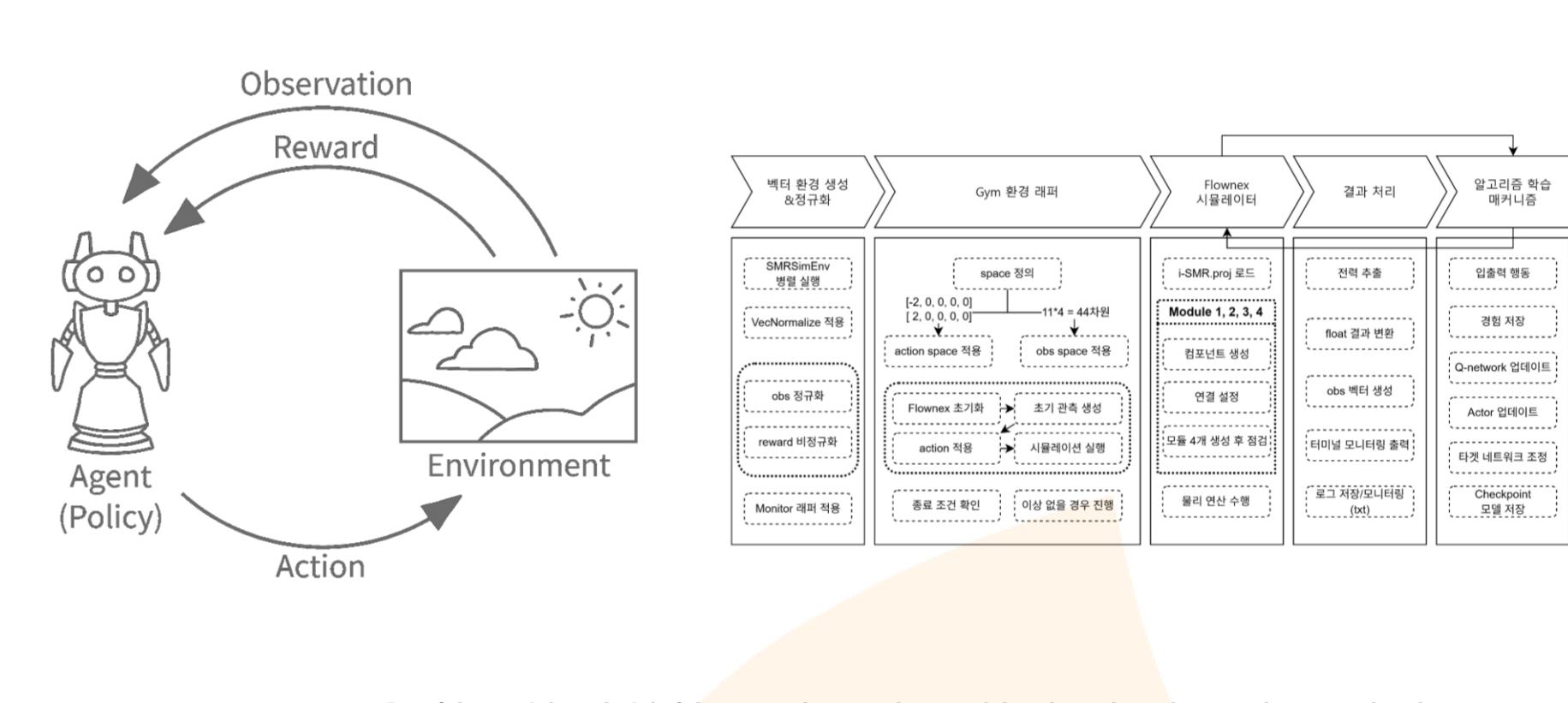
<Flownex SE Nuclear 모델>



<i-SMR Flownex 모델링>



<i-SMR Python API 스크립트>

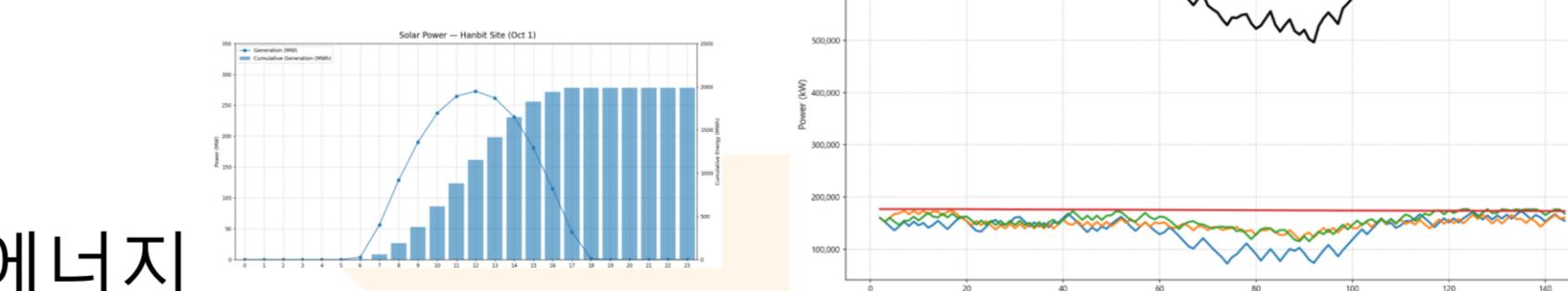


<Gymnasium을 활용한 강화학습 환경 및 알고리즘 설계>

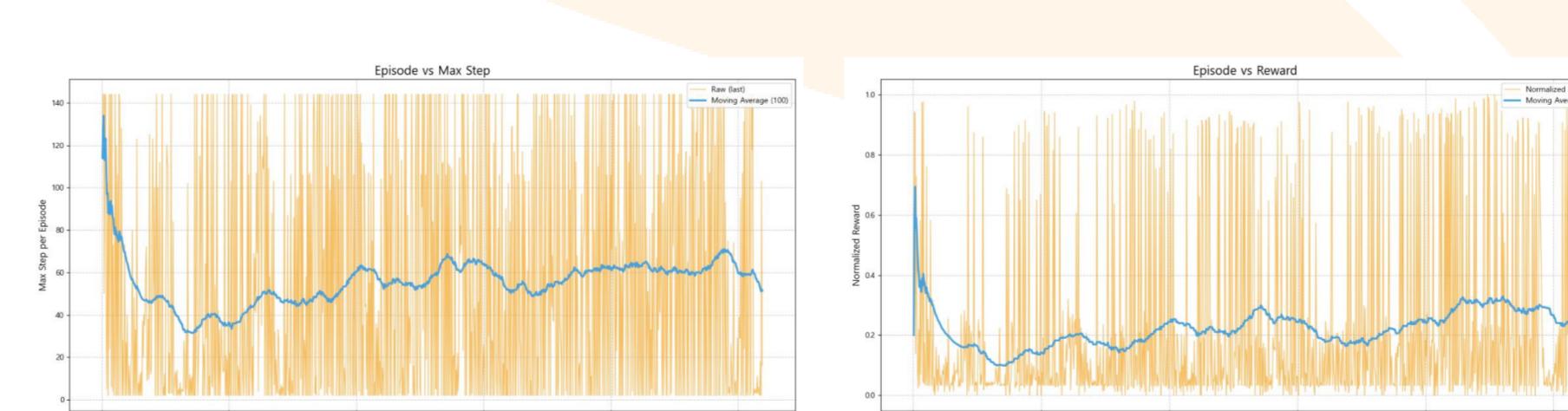
- 강화학습은 명시적인 수학적 모델링 없이도 에이전트가 환경과의 지속적인 상호작용을 통해 누적 보상(Reward)을 최대화하는 최적의 행동 전략, 즉 정책을 활용해 학습하는 머신러닝의 한 분야이며 에이전트는 i-SMR 4개 모듈의 운전을 활용하여 구현
- SAC(Soft Actor-Critic)은 엔트로피(Entropy) 개념을 도입하여 탐험과 활용의 균형을 효과적으로 맞추는 Off-Policy Actor-Critic 계열의 알고리즘
- Gymnasium은 Agent에 Action을 적용하여 Environment의 Observation & Reward를 변환, 벡터 생성&정규화 → Gym 래핑 → 학습 단계를 거치는 구조의 강화학습을 설계
- 알고리즘의 핵심은 Flownex SE의 복잡한 시뮬레이션 데이터를 강화학습이 효율적으로 학습할 수 있는 형태로 변환한 후 개발
- 상태(State)는 한 개의 모듈당 11개의 상태에 대하여 4개 모듈의 주요 변수들을 포함한 총 44개의 연속 변수로 구성, 행동(Action)은 제어봉 삽입도, 메인밸브 개도율, 주급수 펌프 RPM을 기준으로 4개 모듈을 동시에 제어하기 위해 총 12개의 연속 변수로 정의, 보상(Reward)은 에이전트의 학습 방향을 결정하는 보상(Reward)함수는 다수모듈을 고려한 목표를 달성할 수 있도록 항목의 가중 합으로 설계

사례연구

- 부하추종운전을 위한 목표 출력은 기상청 신재생에너지 발전량 시뮬레이션을 활용 네 개 모듈에 대해 초기 연소도를 각각 10%, 50%, 70%, 90%로 설정
- 강화학습 과정에서 에피소드별 보상 변화를 시각화한 결과로, 학습의 수렴 경향과 성능 향상 추세를 파악, 1,000번의 반복 에피소드



<한빛 부지 예측 발전량 및 특정 에피소드 모듈 당 출력 변화>



<에피소드 당 보상 그래프>

$$G_t = \sum_{k=0}^{\infty} r^k r_t + k + 1 \quad J(\pi) = \sum_{t=0}^T E(s_t, a_t) - \rho^{\pi[r(s_t, a_t) + \alpha H(\pi(\cdot|S_t))]}$$

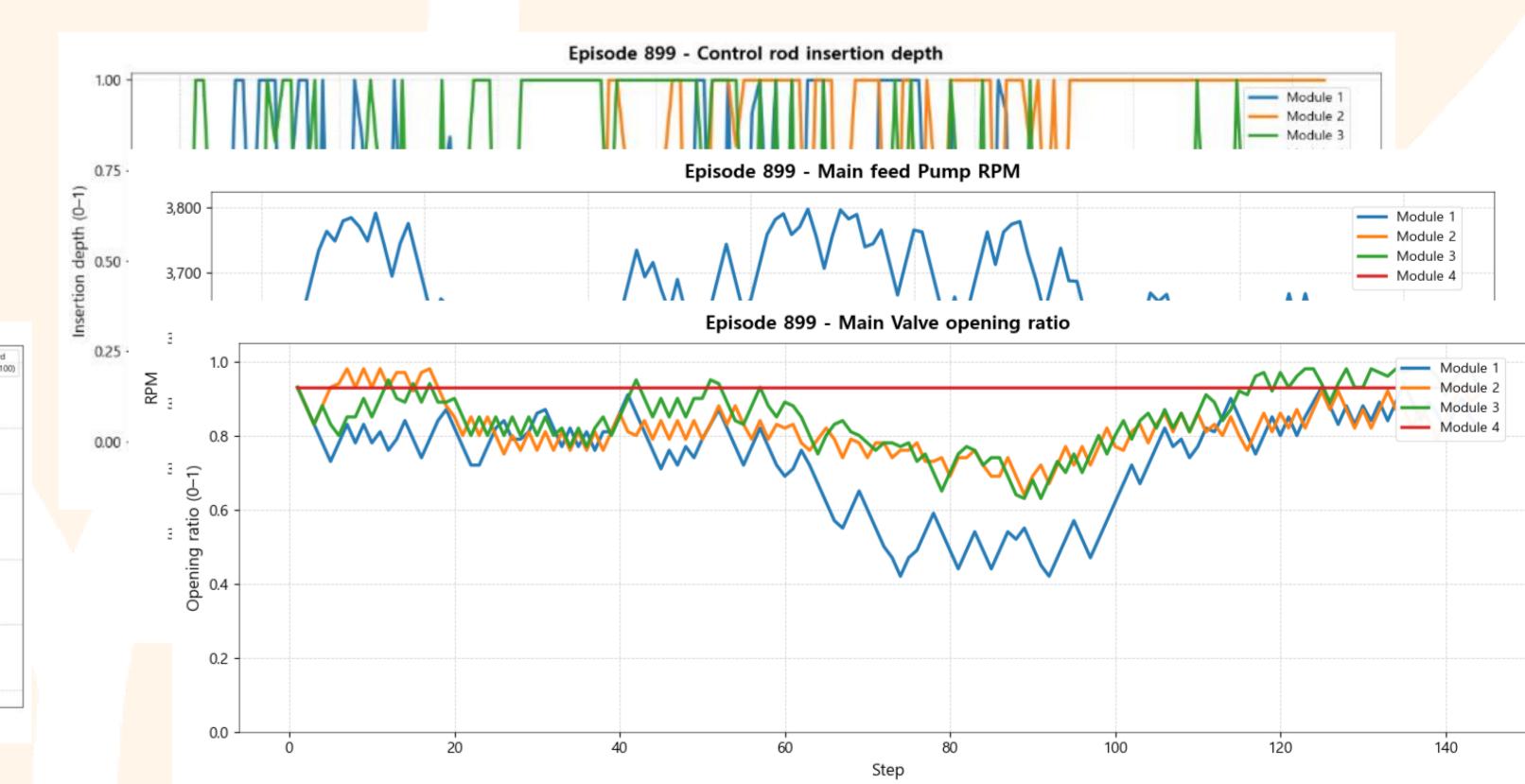
<강화학습 개념>

<SAC(Soft Actor-Critic) 개념>

$$R_t = 10 - 50E_t - \sum_{i=1}^N (a_{i,t} - s_i C_t)^2 - 2000F$$

- 10: 기본 보상 (Base Reward)
- 50E_t: 출력 추종 오차 패널티
- $\sum_{i=1}^N (a_{i,t} - C_t)^2$: 감발량 분배 균형 패널티
- 2000F: 종료(실패) 이벤트 패널티

<강화학습을 위한 SAC 보상 함수>



<부하추종운전에 활용된 액션>

결론 및 향후계획

- 본 연구는 Flownex SE 와 Agent의 정책 설계를 강화학습 알고리즘과 연결시켜주기 위한 Gymnasium 패키지를 사용하였음
- 단일 정책으로 모듈 간 상호작용을 반영한 운전 시나리오를 수행하였으며, 재생에너지 변동 대응 가능성을 검증함. 다만 학습 수렴성과 제어 안정성, 액션 구성의 단순성 등 한계가 존재함
- 향후 P2X(Power-to-X) 연계 및 탄력운전 최적화를 목표로 알고리즘을 개발한다면 활용가치가 상승할 것이라 기대함

제5회 일반대학원 총학생회 공동학술 세미나 발표집