원전의 고온대기 운전 자동화를 위한 강화학습 설계

박재관*, 김택규*, 성승환*, 구서룡*
*한국원자력연구원
Jkpark183@kaeri.re.kr, taekkyukim@kaeri.re.kr, shseong@kaeri.re.kr, srkoo@kaeri.re.kr

A Reinforcement Learning Design for Control Automation in Heat-up Mode of a Nuclear Plant

JaeKwan Park*, TaekKyu Kim*, SeungHwan Seong*, SeoRyong Koo*
*Korea Atomic Energy Research Institute

요 약

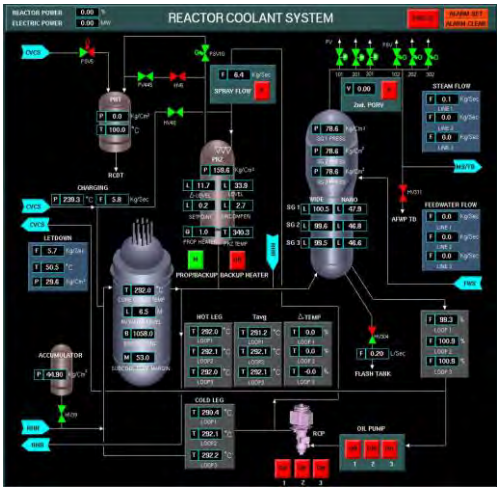
차세대 원전의 계측제어 기술 분야에서는 운영시스템의 자동화 수준을 높이고 운전원의 부담은 낮추기 위한 다양한 연구개발이 진행되고 있다. 최근, 인공지능 기술을 활용하여 원전의 운전에 기여하기 위한 연구가 수행되고 있다. 이 논문은 원전 자동화를 위한 기초 연구로써, 원전 고온대기 모드에서의 자동 제어를 고안하기 위한 강화학습 설계 방법을 소개한다. 기존 원전 시뮬레이터를 강화학습이 가능하도록 확장하였고 강화학습 핵심 요소를 원전 운전에 적합하도록 설계하였다. 실험 결과는 강화학습 기술이 차세대 원전 자동 제어에 적용할 수 있음을 보여준다.

1. 서론

차세대 원전 계측제어시스템 기술은 자동화 수준을 높이고 운전원의 부담을 낮추는 것을 목표로 연구개발되고 있다. 원자력발전소에는 고온대기, 저출력, 전출력, 냉각과 같은 여러 운전 모드가 존재하며 많은 부분이 자동화되어 있으나, 고온대기 운전은 아직도 많은 조치가 수동으로 수행된다. 특히, 고온대기 운전 구간은 발전소 변수를 지속적으로 감시하고 관련된 몇 개의 기기를 빈번히 작동시키는 운전원 조작이 요구된다. 이러한, 잦은 수동 조작은 운전원의 부담을 증가시키고 인적 오류의 가능성을 높이는 문제가 있다. 원전안전운영정보시스템에 공개된 사고 정보에 따르면, 2000 년부터 2019 년까지 원전에서 발생한 발전소 불시정지의 17.8%는 인적 오류에 의한 것으로 보고되었다.

그런데, 원전 상태를 파악하는 것은 패턴 인식과 유사하기 때문에 인공 지능 기술의 적용이 활발히 고려되고 있다. 퍼지 신경망을 이용한 중대사고 상황에서의 의사결정에 대한 연구[1], LSTM (long short-term memory)를 이용한 사고 진단에 관한 연구[2], CNN (convolutional neural network)의 딥러닝 기반 경보 시스템 연구[3] 등 지도학습 기반의 딥러닝 연구가 제안되었다. 최근에는 심층 강화학습[4]에 기반한 비지도 학습이 크게 개선된 성능을 보이고 있다. 특히, Deep Q-Network[5]과 Asynchronous advantage actor-critic[6]은 강화학습의 강력한 성능을 보여주는 계기가 되었다. 이 논문은 원전 제어를 위해서 A3C 강화학습을 설계한 연구 결과를 소개한다.

2. 강화학습을 위한 시뮬레이터 설계

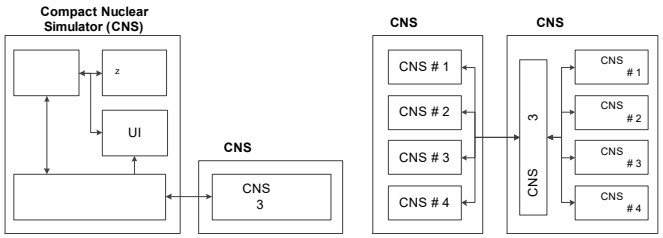


(그림 1) compact nuclear simulator 화면

Compact Nuclear Simulator (CNS)는 발전소의 기초교육, 계통 설계 및 안전성 연구에 필요한 주요 계통을 모의하도록 국내에서 제작된 모의장치이다[7]. CNS 수치 모델은 웨스팅하우스 3-loop 900MW 가압경수로형 발전소를 모의할 수 있다. CNS 운전은 저온 정지, 고온 정지, 고온 대기, 저출력, 전출력 등의 운전 모드로 구분된다. 이 CNS 는 이 논문에서 활용할 시뮬레이터로써 선정되었으나, Stand-alone 모드만으로 동

작하며 데이터 통신이 지원되지 않기 때문에 강화학습 목적에 맞도록 변경이 요구된다. 그림 1 은 CNS 의 주요 변수를 실시간으로 확인 할 수 있는 인터페이스 화면이다.

준비된 데이터를 사용하는 지도 학습과 달리, 강화학습은 시뮬레이터와 실시간으로 연결되어야 하고 플랜트의 상태를 스텝(step) 단위로 획득할 수 있어야 한다. 이를 위해, 이 논문에서는 기존 CNS 구조를 그림 2 와 같이 확장하였다. 먼저, 서버 측에서는 서버의 신호 정보를 클라이언트로 전송할 수 있는 통신 모듈이 추가되었고 클라이언트 측에서는 신호 정보를 수신할 수 있는 통신 모듈이 추가되었다. 또한, 클라이언트의 제어 신호를 서버로 전달하는 기능, 클라이언트가 서버의 상태 정보를 스텝 단위로 수신할 수 있도록 해주는 기능을 추가하였다. 뿐만 아니라, 다중 클라이언트가 병렬 학습을 수행할 수 있도록 다중 CNS 실행 환경을 구축하였다.

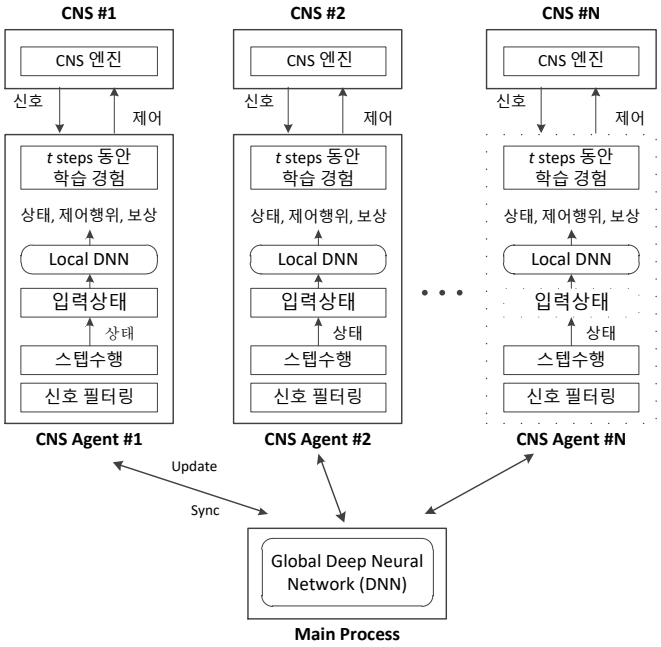


(그림 2) 다중 에이전트를 위한 CNS 커스터마이징

3. 강화학습 핵심 항목 설계

강화학습을 수행하기 위해서는 몇 가지 핵심 항목이 설계되어야 한다. 먼저, CNS 운전을 학습하는 에이전트(agent), 에이전트가 상호작용하는 환경(environment), 에이전트에 의해 인지되는 원전 상태(state), 에이전트에 의해 수행되는 제어 행위(action), 그리고 에이전트의 제어 행위에 따른 결과로 받게되는 보상(reward)이 설계되어야 한다. 특히, 본 논문의 CNS 환경은 CNS 운전을 시작하거나 종료하는 역할을 수행하고 에이전트가 선택한 제어 행위를 실제로 수행하는 역할과 제어 결과에 의한 변경된 CNS 상태를 제공하는 역할을 수행한다.

환경 설계에서는 다양한 상황에 대한 동시다발적인 학습이 중요하므로, 매우 긴 운전과정을 11 개 부분으로 나누어 각각을 시나리오의 시작 지점을 정의하였다. 즉, 하나의 시나리오는 하나의 학습 에피소드(episode)가 되고 각 에이전트는 임의의 시나리오로 진입한다. 시나리오 설계에서 중요했던 부분은 각 시나리오의 종료와 다음 시나리오의 시작이 맞닿도록 구성하면 각 시나리오 시작지점에 대한 학습이 빈약해지는 현상이 발생한다. 이를 보완하기 위해서, 각 시나리오가 절반씩 겹치도록 구성하였다. 그리고, 설계된 환경에서는 강화학습에서 에이전트가 딥러닝을 통해 출력하는 행위 번호 (action number)를 실제 CNS 의 제어 명령으로 치환하고 스텝 단위로 원전 상태 변화를 제어한다. 또한, 설계된 환경은 원전 상태 신호들을 에이전트의 입력 형태로 변환해주도록 설계되었다.



(그림 3) 다중 에이전트의 학습 구조

기본적으로 에이전트는 강화학습의 목표를 달성하기 위해서 환경의 현재 상태를 관찰하고, 제어 행위을 선택하고, 피드백되는 보상에 따라 학습을 수행한다. 에이전트가 학습을 수행할 때, 동시에 다양한 상태를 학습하는 것이 중요한데, A3C[6]는 다중 쓰레드 개념으로 그 문제를 해결한다. 이 논문은 A3C[6]의 개념을 활용하여 다중 CNS 에 다중 쓰레드를 연결하므로 병렬 학습이 가능한 구조로 에이전트를 설계하였다.

그림 3 에서 보이는 것과 같이, 메인 프로세스에 의해 각 에이전트의 태스크는 쓰레드로 생성되고 Global DNN (Deep Neural Network)의 weights 는 각 에이전트에 동기화 된다. 에이전트는 CNS 엔진으로 받은 신호들을 필터링 하여 강화학습에 적합한 상태 구조로 변환한다. DNN 을 이용하여 제어 행위를 선정하고 이것을 CNS 제어 명령으로 변환한 후 CNS 엔진에 전송하면 명령이 실행되고 결과에 따른 보상을 획득한다. 이러한 실행 경험은 여러 스텝 동안 저장하고 주기적으로 Global DNN 을 업데이트 하며, 업데이트 후에는 다시 그 global DNN 을 에이전트로 동기화

한다.

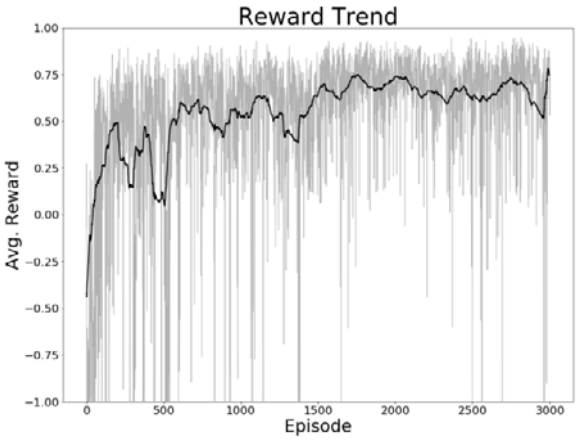
강화학습의 상태 (state)는 2 차원 배열로 설계하였다. CNS 상태를 대표하는 신호들을 선정하여 현재 상태 (current state)의 첫 번째 차원에 배치한다. 연관성이 높은 신호들은 DNN 이 합성곱 (Convolution)을 수행할 때 고려되도록 인접하여 배치하였다. 각 신호 값은 두 번째 차원에 맵핑한다. 스텝 마다 얻는 상태(state)를 누적하여 2 차원 배열을 쌓으면, 3 차원 배열이 되고, 이것을 강화학습에서 사용하는 DNN 의 입력 상태(input state)로 사용한다. 이를 기반으로, 이 논문에서 사용하는 입력 상태는 (20×100×8)로 설계하였다.

강화학습의 행위(Action)는 고온대기 운전에서 조작이 필요한 가압기 스프레이밸브(Spray valve)의 열림(open)과 닫힘(close), 가압기 출구 측 유량 조절 밸브의 열림(open)과 닫힘(close)에 대한 인덱스 번호를 붙인 형태로 표현하였다. 에이전트는 매 스텝 마다 입력 상태에 대하여 이러한 제어 행위 중에서 한 가지를 선택하여 수행하게 된다.

강화학습에서 중요한 설계요소 중 하나가 보상(Reward)이다. 특히, 빈도가 드문 보상(sparse reward)나 지연된 보상(delayed reward)로 설계하면 학습이 잘되지 않는 현상이 발생한다. 이 환경에서는 각 스텝 마다 규칙에 따른 보상을 받도록 설계하였다. 먼저, 가압기 압력(Pressurizer pressure) 규칙을 위해서 냉각재 계통 온도(RCS average temperature) 대비 가압기 압력의 직선식을 정의하고 그 직선식으로부터의 거리 가가까울수록 높은 보상을 멀어질수록 낮은 보상이나 부정적 보상(negative reward)을 부여한다. 가압기 수위(Level) 규칙은 항상 일정한 범위 내에서 유지하도록 수위 범위의 중간값으로부터의 거리를 기준으로 보상을 부여한다. 즉, 주요 상태가 운전 범위 내에 잘 유지될 경우 높은 보상을 받게 된다. 각 스텝의 상태에 대해서 두 가지의 보상을 합친 값이 최종 보상으로 반영되도록 설계하였다.

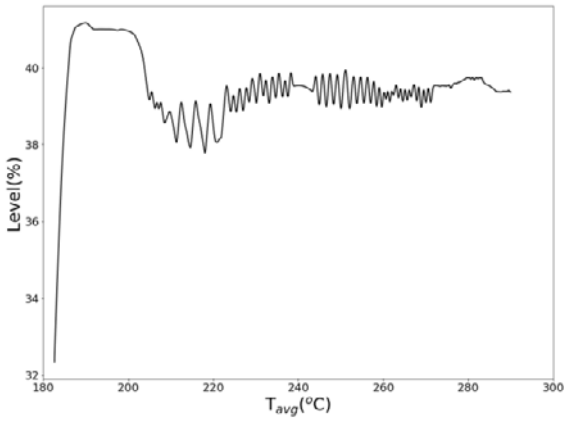
4. 학습 및 실험 결과

실험은 CNS 의 약 180 ℃의 고온정지에서 약 290 ℃의 고온대기까지의 운전을 수행하였다. 시뮬레이터인 CNS 는 인텔 코어 i7 CPU 의 Linux OS 가 설치된 여러대의 컴퓨터에서 서버 소프트웨어를 다중으로 실행하고 1 대의 컴퓨터에서 클라이언트 소프트웨어를 실행한다. 강화학습은 클라이언트 컴퓨터에서 실행되고 4 개의 쓰레드가 동시에 CNS 서버 소프트웨어에 연결되어 학습을 수행하는 환경을 설정하였다. 또한, DNN 을 위해서 케라스 딥러닝 프레임워크를 활용하였다.



(그림 4) 학습과정에서의 보상값 추세

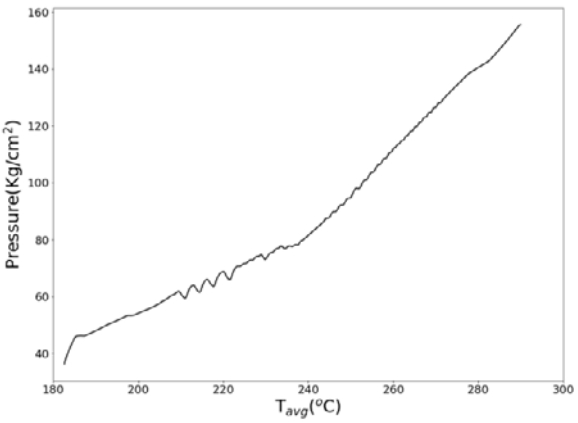
그림 4 는 학습과정에서 각 시나리오에서 얻을 수 있는 최대 보상을 1 로 정의하고, 학습동안에 각 스텝 마다 얻은 보상을 나타낸 그래프이다. 각 스텝 마다 최대 0.05 를 얻을 수 있으나 각 시나리오 에피소드 마다 최대 스텝 수가 다르기 때문에 최대 획득 가능 보상 도 다르다. 초반 학습에서는 원자로 정지가 발생하거나 부정적 보상을 많이 받으면서 전체 보상이 낮게 나오는 결과를 보였다. 그리고, 각 시나리오의 시작 상태가 보상이 낮은 상태에서 시작하기 때문에 최적 상태까지 이동하는 동안에 낮은 보상을 계속 받는다. 또한 학습과정에서는 더 나은 선택을 탐색하기 위해서 랜덤하게 행위를 선택할 확률을 계속 유지한다. 이러한 특징 때문에 실제 학습된 결과 보다는 낮게 나왔지만, 점차 학습을 거치면서 획득 보상이 증가하는 경향을 보였다.



(그림 5) 실험결과 - 가압기 수위 제어

그림 5 는 학습이 종료된 결과물로 시험 제어를 수행한 결과 중에 가압기 수위가 제어된 결과를 보여준

다. 초기에는 매우 낮은 수위를 빠르게 보충하기 위해서 살수 밸브를 열림을 크게 하였고, 이후 빠르게 큰 수위 변화와 작은 수위 변화에 잘 대처하는 제어 경향을 보였다. 마지막 부분에는 점차 수위 변화가 안정되고 제어의 범위도 좁혀지는 경향을 보였다. 수위는 적정 범위인 30~50% 이내에서 잘 유지되었다.



(그림 6) 실험결과 - 가압기 압력 제어

그림 6 은 시험 제어를 수행한 결과로써 가압기 압력의 변화를 보여주는데, 매우 좋은 결과를 보였다. 실제로 운전원이 운전하는 경향과 유사한 형태의 결과를 보였으며, CNS 저출력 운전에서 최소한 고려되어야 하는 P-T(pressure-temperature) 커브의 최소와 최대 압력 경계 (minimum and maximum pressure boundaries) 이내에서 잘 운전되었다.

5. 결론

수동으로 감시 및 제어되는 원전의 고온대기 모드에서의 운전은 자동화를 위해 많은 연구가 필요한 부분이다. 이 논문은 차세대 원전의 자동화를 위한 기초 연구로써, 고온대기 모드에서의 자동 제어를 위한 강화학습 설계 방법을 소개하였다. 실험 결과에서는 고도화를 거쳐 활용가능성이 높음을 확인하였다. 이러한 결과를 바탕으로, 향후 더 복잡한 감시 및 제어가 요구되는 분야에 적용성을 분석해볼 계획이다.

Acknowledgement

이 논문은 2019 년 과학기술정보통신부의 재원으로 한국연구재단의 지원을 받아 수행된 연구임(No. 2019M2C9A1055903).

참고문헌

[1] Yoo K.H., Back J.H., Na M.G., Hur S., Kim H., “Smart support system for diagnosing severe accidents in nuclear power plants”, Nuclear Engineering and Technology, Vol.50, pp.562-569, 2018

[2] Yang J., Kim J., “An accident diagnosis algorithm using long short-term memory”, Nuclear Engineering and Technology, Vol.50, pp.582-588, 2018

[3] Kim T.K., Park J.K., Lee B.H., Seong S.H., “Deep-learning-based alarm system for accident diagnosis and reactor state classification with probability value”, Annals of Nuclear Energy, Vol.133, pp.723-731, 2019

[4] Mnih V., Kavukcuoglu K., Silver D., Rusu A.A., Veness J., Bellemare M.G., Graves A., Riedmiller M., Fidjeland A.K., Ostrovski G., Petersen S., Beattie C., Sadik A., Antonoglou I., King H., Kumaran D., Wierstra D., Legg S., Hassabis D., “Human-level control through deep reinforcement learning”, Nature, Vol.518, pp.529-533, 2015

[5] Silver D., Huang A., Maddison C.J., Guez A., Sifre L., Van Den Driessche G., Schrittwieser J., Antonoglou I., Panneershelvam V., Lanctot M., Dieleman S., Grewe D., Nham J., Kalchbrenner N., Sutskever I., Lillicrap T., Leach M., Kavukcuoglu K., Graepel T., Hassabis D., “Mastering the game of Go with deep neural networks and tree search”, Nature, Vol.529, pp.484-489, 2016

[6] Mnih V., Badia, A.P., Mirza M., Graves A., Lillicrap T., Harley T., Silver D., Kavukcuoglu K., “Asynchronous methods for deep reinforcement learning”, 33rd International conference on machine learning, New York, USA, 2016

[7] Kwon K.C, Park J.C., Jung C.H., Lee J.S., Kim J.Y., “Compact nuclear simulator and its upgrade plan”, Korea Atomic Energy Research Institute, 1997.