스마트 팩토리 환경에서 설비 고장 복구를 위한 심층 강화학습 기반 로봇 스케줄링 시스템에 관한 연구

이승준, 이유현, 하승호, Asadbek Khalilov, 유상조

인하대학교

pank147@inha.edu, dldbgus203@inha.edu, h04656@inha.edu, khalilovasadbek@inha.edu, sjyoo@inha.ac.kr

A Deep Reinforcement Learning-Based Robotic Scheduling System for Equipment Failure Recovery in Smart Factories

Seung-Joon Lee, Yu-Hyun Lee, Seung-Ho Ha, Asadbek Khalilov, Sang-jo Yoo

Inha Univ.

요 약

본 논문은 스마트 팩토리 환경에서 설비 고장이 빈번하게 발생하는 상황에서, 로봇의 효율적인 의사결정을 통해 생산성을 유지·개선 할 수 있는 지능형 제어시스템을 제안한다. 이를 위해 심층 강화학습의 일종인 심층 Q 네트워크(DQN, Deep Q Network)를 활용한 의사결정 프레임워크를 구성하였다. 제안된 시스템은 공정 내 설비 상태와 생산 흐름을 반영한 상태 정보를 기반으로 로봇의 행동을 결정하도록 설계하였으며, 설비 고장 상황에서 생산성이 향상됨을 확인하였다.

Ⅰ. 서 론

본 논문에서는 스마트 팩토리 환경에서 설비 고장이 발생하는 상황을 대상으로, 공정 내 로봇의 의사결정이 생산성에 미치는 영향을 분석하고 이를 개선하기 위한 자동화 시스템을 제안한다. 스마트 팩토리에서는 자동화 로봇이 설비 점검, 공정 복구, 병목 완화와 같은 역할을 수행하고 있다. 그러나 기존의 사전 정의된 로봇 제어 방식은 고정된 의사결정 규칙에 기반하고 있어, 설비 고장이나 병목과 같은 동적 상황에 유연하게 대응하지 못하는 한계를 지닌다.[1] 이에 본 연구는 심층 강화학습의 일종인 심층 Q 네트워크 (DQN, Deep Q-Network)를 활용하여 설비 고장 상황에서 공정 상태와 생산 흐름을 고려하여 로봇 의사결정을 수행할 수 있는 지능형 의사결정 시스템을 제안한다. [3] 본 연구의 목표는 심층 강화학습 기반 로봇 스케줄링을 통해 설비 고장 상황에서도 생산성을 유지·개선할 수 있는 로봇 스케줄링 기법을 제시하는 데 있다.

Ⅱ. 심층 강화학습을 이용한 설비 고장 복구를 위한 로봇 의사 결정 방법 제안

본 장에서는 심층 강화학습을 이용한 공장 설비 고장 복구를 위해, DQN 알고리즘을 이용한 방법론을 제안한다. 전체 시스템 아키텍처는 그림 1 과 같이 구성한다.

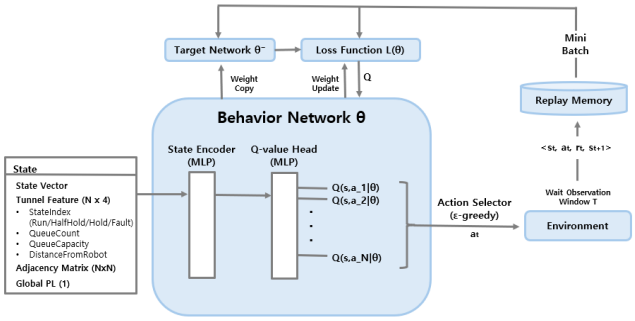

그림 1. 제안하는 로봇 스케줄링 모델의 구조

2-1. 스마트 팩토리 환경 및 설비 고장 복구 문제 정의

본 연구에서는 스마트 팩토리 환경에서 발생하는 설비 고장 상황을 고려한 로봇 기반 설비 복구 문제를 다룬다. 실험 환경은 총 N 개의 설비 노드로 구성된 공장의 라인을 의미하며, 각 설비는 네트워크 형태로 연결되어 제품이 순차적으로 이동하는 구조를 가진다. 공장의 작동 중 설비 노드의 고장이 발생하면 해당 설비 노드는 정상적인 공정의 역할을 수행할 수 없으며, 이로 인해 생산율 저하 및 대기열 증가와 같은 문제가 발생한다. 이동 로봇은 고장난 설비 노드 중 하나를 선택하여 해당 위치로 이동한 뒤 수리를 수행한다.

2-2. Deep Q-Network 기반 로봇 스케줄링 시스템

본 연구에서는 설비 고장 복구 문제를 해결하기 위해 DQN 기반의 로봇 의사 결정 시스템을 제안한다. 그림 1 은 제안하는 DQN 기반 로봇 의사 결정 시스템의 전체 구조를 나타낸다.

환경(Environment)은 스마트 팩토리 환경으로, 각 시점에서 공장 전체의 상태를 상태 벡터 s_t로 제공한다. 상태에는 공장 운영에 영향을 미치는 주요 지표들이 포함된다.

Q 값은 식(1)과 같이 정의된다. r_t는 시점 t에서 수행한 설비 노드 수리 행동에 대한 보상이며, γ는 미래 보상의 중요도를 조절하는 할인 인자이다. 본 연구에서는 정책 네트워크(Policy Network)와 타깃 네트워크(Target Network)를 분리한 Double DQN 구조를 사용함으로써, Q 값의 과대 추정 문제를 완화하고 학습의 안정성을 향상시킨다.

$$Q(s_t, a_t) = r_t + \gamma Q_{target}(s_{t+1}, argmax_a Q_{policy}(s_{t+1}, a)) \quad (1)$$

로봇이 한 step 동안 수행한 경험은 $<s_t, a_t, r_t, s_{t+1}>$ 형태의 튜플로 Replay Memory 에 저장된다. 이후 무작위로 추출된 미니 배치를 사용하여 경사하강법 기반 학습을 수행하여, 상태 간 상관성을 줄이고 효율적인 학습이 가능하도록 한다.

손실 함수는 식(2)와 같이 정의된다. M 은 미니 배치의 크기이며, $a_j^* = argmax_a Q_{policy}(s_{j+1}, a)$ 가 된다. 타깃 네트워크는 주기적 또는 소프트 업데이트 방식을 통해 정책 네트워크의 파라미터를 반영하도록 설계된다.

$$Loss = \frac{1}{M}\sum_{j=1}^{M}(r_j + \gamma Q_{target}(s_{j+1}, a_j^*) - Q_{policy}(s_j, a_j))^2 \quad (2)$$

2-3. Proposed DQN Network 구조

그림 2 는 본 논문에서 제안하는 DQN Network 의 전체 구조를 나타낸다. 공장 상태는 State Vector 형태로 입력되며, 각 계층으로 전달되어 상태-행동 쌍에 대한 가치 함수로 변환된다. 이 과정에서 공장의 환경적 요소와 로봇의 수리에 따른 노드 간 관계가 통합적으로 학습이 이루어진다. 최종적으로 네트워크는 최대 Q 값을 갖는 행동을 선택하며, 의사결정 행동 공간은 총 N 개의 조작으로 구성된다.

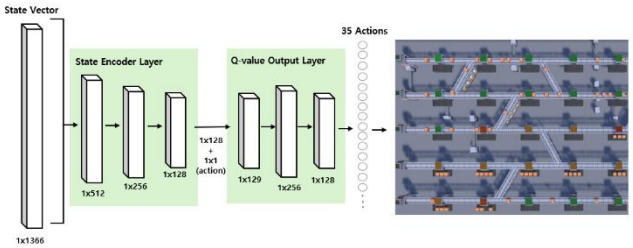

그림 2. Proposed DQN Network 구조

2-4. 상태, 행동, 보상 정의

로봇 의사 결정 시스템에서의 상태(state), 행동(action), 보상(reward)은 다음과 같이 정의된다.

- 상태(state): 본 연구에서는 공장 전체의 운영 상황을 하나의 고정 길이 벡터로 표현한다. 상태 벡터는 각 노드의 동작 상태, 큐 정보, 로봇과의 거리와 같은 노드 단위의 국소 정보와, 노드 간 연결 관계를 나타내는 인접 행렬, 그리고 관찰 구간 동안의 전역 생산성 지표로 구성된다.

- 행동(action): 행동은 로봇이 수리 대상으로 선택하는 공장 내 노드를 의미한다. 즉, 행동은 공장에 존재하는 노드 집합 중 하나를 선택하는 이산 행동으로 정의되며, 강화학습 모델은 각 후보 노드에 대해 Q-value 를 평가한 후 가장 적절한 노드를 선택한다.

- 보상(reward): 본 연구의 목표는 고장이 발생한 공장 환경에서 전반적인 생산성을 효과적으로 유지 및 향상시키는 것이다. 이를 위해, 본 연구에서는 공정 전체의 운영 효율을 반영할 수 있도록 보상 함수를 설계하였다.[2] 총 보상 R^{total}은 시간 구간 T 동안의 생산율($\overline{PL}(T)$), 고장 또는 병목으로 인해 공장 내에 누적되는 대기 제품 수($\overline{QS}(T)$), 그리고 고장난 설비 노드가 인접한 설비 노드의 생산 흐름에 미치는 영향을 반영하는 항의 합($\overline{RB}(T)$)으로 구성된다.

$$R^{total} = w_1\overline{PL}(T) - w_2\overline{QS}(T) - w_3\overline{RB}(T) \quad (3)$$

Ⅲ. 모의 실험 결과

제안한 심층 강화학습 기반 로봇 스케줄링 기법의 성능을 검증하기 위해, Unity 엔진을 활용한 스마트 팩토리 가상환경을 구축하고 모의 실험을 수행하였다. 본 가상환경은 설비 노드 간 연결 구조, 제품의 이동 흐름, 설비 고장 및 복구 과정을 포함하도록 구현되었다. 실험 결과로 얻어진 보상 값의 합은 그림 3 을 통해 나타내어 모델의 성능을 분석하였다.

Parameter	Value	Parameter	Value
Total Node (N)	35	Target Network Update	1000
Learning Rate	3×10^{-4}	Replay Memory Capacity	100000
Discount Factor γ	0.99	Batch Size	64

표 1. 시뮬레이션 파라미터

제안한 방법의 성능 비교를 위해, 무작위 설비 노드의 고장 시나리오 하에서 학습 단계에 따른 로봇 스케줄링 전략의 변화를 통해 분석하였다. 각 시나리오마다 $k = 0{\sim}3$ 개의 설비 노드가 무작위로 고장 나고, 로봇은 고장난 설비 노드 집합 내에서 수리 대상을 선택한다. 이후 1 episode = n scenarios로 설정하여 에피소드 당 보상을 계산한다.

학습 초기에는 ε-*greedy* 탐험 전략에 의해 탐험 비중이 높아 로봇의 선택이 탐험 중심으로 이루어지며, 이에 따라 설비 수리 순서 또한 일정한 패턴을 보이지 않는다. 반면, 학습이 진행됨에 따라 ε가 감소하며 Q 값 기반의 선택 비중이 증가하고, 공장 상태를 고려한 일관된 설비 수리 순서가 형성된다.

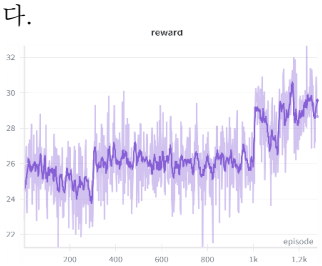

그림 3. 학습 결과

그림 3 은 학습 에피소드에 따른 평균 누적 보상의 변화를 나타낸다. 학습 초기 구간에서는 탐험 중심의 행동 선택으로 인해 보상 값의 변동성이 크게 나타나지만, 학습이 진행됨에 따라 평균적인 보상 값이 점진적으로 증가하는 경향을 보인다. 이는 로봇이 반복적인 상호작용을 통해 설비 노드 고장 상황에서 공장 전반의 생산성을 고려한 수리 우선순위를 학습하고 있음을 의미한다.

Ⅳ. 결론

본 논문에서는 심층 강화학습(DQN) 기반의 로봇 의사결정 시스템을 구현하여, 실시간으로 수리 대상을 선택하고, 공정 흐름을 효율적으로 유지할 수 있는 시스템을 제안한다. 실험 결과, 제안된 강화학습 기반 접근 방식은 기존의 규칙 기반 방식에 비해 공정 정체 상황을 보다 효과적으로 완화하고, 전반적인 생산 흐름 측면에서 안정적인 성능을 유지하였다. 이러한 결과는 심층 강화학습 기반 로봇 의사결정 기법이 복잡한 공장 환경에서도 전역적인 성능을 고려한 자율적 판단이 가능함을 입증하며, 향후 지능형 공장 및 자율 운영 시스템으로의 확장 가능성을 보여준다.

ACKNOWLEDGEMENT

This work was supported by the IITP(Institute of Information & Communications Technology Planning & Evaluation)-ITRC(Information Technology Research Center) grant funded by the Korea government(Ministry of Science and ICT) (RS-2021-II212052).This work was supported by the National Research Foundation of Korea(NRF) grant funded by the Korea government(MSIT) (RS-2025-21212972)

참 고 문 헌

[1] Y. Zhang, L. Wang and L. Gao, “Smart manufacturing scheduling : A literature review, ” Journal of Manufacturing Systems, vol. 61, pp. 265-271, 2021

[2] Z. T. Zhou, D. Tang, H. Zhu, and Z. Zhang, "Reinforcement learning with composite rewards for production scheduling in a smart factory," IEEE Access, vol. 9, pp. 752-766, 2021.

[3] L.Zhou, L. Zhang, and B. K. P. Horn, “Deep reinforcement learning-based dynamic scheduling in smart manufacturing,” Procedia CIRP, vol. 93, pp. 383 - 388, 2020.