

## 강화학습 기반의 AGV 스케줄링

이세훈\*, 김재승<sup>o</sup>, 염대훈\*, 문환복\*, 이창호\*\*

<sup>o</sup>인하공업전문대학 컴퓨터시스템과

\*\*한국콘베어공업(주) 기술연구소

e-mail: seihoon@inhatc.ac.kr\*, jsman9777@gmail.com<sup>o</sup>, {ydh950901, casio1994}@naver.com\*, changholee@conveyor.co.kr\*\*

## Reinforcement Learning based AGV Scheduling

Se-Hoon Lee\*, Jea-Seung Kim<sup>o</sup>, Dae-Hoon Yeom\*, Hwan-Bok Mun\*, Chang-Ho Lee\*\*

<sup>o</sup>Dept. of Computer System, Inha Technical College

\*\*R&D Center, Korea Conveyor Ind. Co.,Ltd.

### ● 요약 ●

스마트 팩토리의 핵심 요소 중 하나인 AGV를 운용하기 위해서 스케줄링은 간과할 수 없는 문제이다. 기존의 정적인 휴리스틱 방식은 실시간으로 운용되는 스마트 팩토리에 다소 부적합한 면이 있다. 본 논문에서는 이러한 스케줄링에 관한 문제를 해결하고자 SLAM 기반의 자율주행 AGV를 운용 할 수 있는 3D 가상 환경을 설계하고 해당 환경에서 강화학습을 기반으로 한 스케줄링을 구현해 실시간으로 변화하는 공장에 적합한 동적인 스케줄링을 설계하였다.

**키워드:** 스케줄링(Scheduling), AGV(Automated Guided Vehicle), 강화학습(Reinforcement Learning)

### I. Introduction

많은 나라와 기업들이 지능형 공장을 만들고자 힘을 쏟고 있다. 아마존에서는 물류 운반용 AGV(Automated Guided Vehicle)가 사람을 대신해 물건을 운반 하고 있다. AGV를 운용하는데 있어 자율주행, SLAM의 기술들과 스케줄링이 중요한 문제로 인식된다[1]. AGV 스케줄링에는 유전자, SA와 같은 다양한 알고리즘이 있는데 이 방식은 정적으로 스케줄링을 수행하는 방식이다. 정적인 스케줄링 방식은 공장과 같이 실시간으로 바뀌는 환경에는 문제가 될 수 있다.

본 논문에서는 스케줄링 문제를 해결하고자 Unity를 통해 AGV와 공장 환경을 가상 환경에 구현하였고, DQN(Deep Q-Network) 기반의 강화학습을 통해 정적인 스케줄링이 아닌 환경에 따라 변화하는 동적인 스케줄링을 도입하였다[2].

### II. Design of the System

#### 1. System Architecture

Fig. 1은 강화학습 기반의 AGV 시스템 구조도이다.

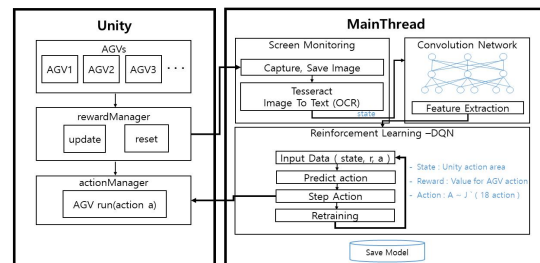


Fig. 1. System Architecture

Unity에서는 스마트팩토리 환경을 다중 에이전트로 구성하였고, 학습에 사용 될 보상은 rewardManager에서 시나리오에 따라 보상을 update, reset 으로 업데이트한다. Main에서는 OCR로 가상 환경의 action과 총 reward를 확인한다. 가상환경에서 2D로 구성된 공장 이미지를 CNN를 통해 피처를 추출하게 된다. 추출한 피처와 총 reward, action등을 DQN의 input data로 학습을 진행한다. 예측한 액션을 가상 환경에 전달하게 되며, 액션이 실행된 후 누적된 reward, state, action을 기반으로 반복 학습을 진행하게 된다.

Fig. 2는 스마트팩토리 화면이며, Fig 3은 강화학습 스케줄링 시스템의 흐름도이다.

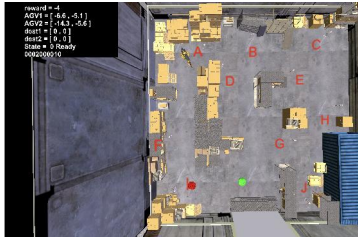


Fig. 2. Unity Virtual Environment

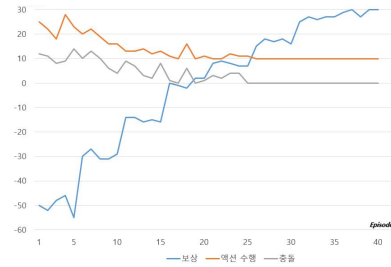


Fig. 4. Experiments Graph

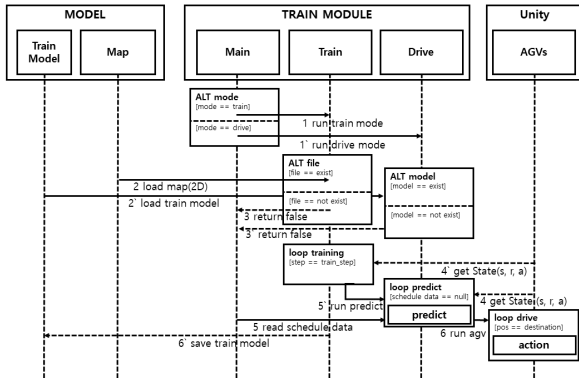


Fig. 3. Sequence Diagram

### III. Conclusions

본 논문에서는 AGV의 정적인 스케줄링 문제를 해결하기 위해 강화 학습 기반의 동적 스케줄링 시스템을 도입하였다. Unity로 공장 환경과 AGV를 가상으로 구현하였고, AGV의 동작 시나리오와 스케줄러를 작성하고 DQN 강화학습으로 최적화된 스케줄표를 도출했고 도출된 표를 토대로 AGV를 운용하였다. 그 결과 기방문 노드에 방문하는 횟수가 감소하고, 가까운 AGV 동작, AGV 간의 충돌 횟수 감소 등의 결과를 확인할 수 있었다. 이 시스템을 통해 기존의 정적인 시스템을 벗어나 동적인 환경에서 스케줄링 문제에 강화학습이 적용 가능성을 보였다.

크게 학습 모델과 맵이 저장될 MODEL 부와 학습모델과 주행모듈이 포함된 TRAIN MODULE 부와 가상환경 부로 나누어져 있다. 학습이 시작되면 맵을 로드하게 된다. 로드된 맵과 가상환경에서의 상태(state, reward, action)을 불러와 학습을 진행하게 된다. 스텝마다 예측을 하며 AGV를 동작시키게 되고 재설정된 값을 토대로 반복 학습을 진행한다. 주행모드에서는 학습된 모델을 로드하게 된다. 스케줄 데이터가 있으면 반복적으로 예측을 하게 되고 예측된 값을 토대로 Unity에서는 AGV가 동작하게 된다.

### 2. Experiments

본 시스템의 시나리오는 다음과 같다.

- agent : Multi AGV (2)
- node : 10 nodes ( A~J )
- action : 20 actions ( a, a', b, ..., j')
- reward : Collision detection ( reward = -2 )
  - Nearer agv run ( reward = 3 )
  - Node arrival ( reward = 1 )
  - Revisit node ( reward = -2 )

테스트 결과 Fig. 4와 같이 에피소드가 거듭 될수록 스케줄러 성능의 향상을 확인할 수 있었다. 충돌 없이 모든 노드를 방문한 뒤 에피소드가 종료되는 경우 해당 에피소드의 총 액션 수, 방문한 노드의 수가 줄어드는 것을 확인할 수 있다. 즉, 기방문한 노드임을 확인하고 방문하지 않은 노드를 방문하려 하는 것을 확인할 수 있었다. 또한 충돌 횟수조차 감소하였다. 방문 예정인 포인트에 다른 AGV가 이미 위치해 있다면 해당 노드는 방문하지 않으려는 경향을 확인할 수 있었다. 실험 결과 전체적인 보상의 증가를 확인 하였고 이는 곧 제안하는 시스템의 성능을 의미한다.

### REFERENCES

[1] Maryam Mousavi, Hwa Jen Yap\*, Siti Nurmaya Musa, Farzad Tahriri, Siti Zawiah Md Dawal. Multi-objective AGV scheduling in an FMS using a hybrid of genetic algorithm and particle swarm optimization

[2] To appear, AAAI 2018. Cellular Network Traffic Scheduling with Deep Reinforcement Learning