ISSN 1598-4095 (Print)
ISSN 2287-7401 (Online)
http://dx.doi.org/10.9712/KASS.2023.23.2.29

Journal of Korean Association for Spatial Structures
Vol. 23, No. 2 (통권 92호), pp.29~36, June, 2023

다중 에이전트 강화학습을 이용한 RC보 최적설계 기술개발

Development of Optimal Design Technique of RC Beam using Multi-Agent Reinforcement Learning

강 주 원*
Kang, Joo-Won

김 현 수**
Kim, Hyun-Su

Abstract

Reinforcement learning (RL) is widely applied to various engineering fields. Especially, RL has shown successful performance for control problems, such as vehicles, robotics, and active structural control system. However, little research on application of RL to optimal structural design has conducted to date. In this study, the possibility of application of RL to structural design of reinforced concrete (RC) beam was investigated. The example of RC beam structural design problem introduced in previous study was used for comparative study. Deep q-network (DQN) is a famous RL algorithm presenting good performance in the discrete action space and thus it was used in this study. The action of DQN agent is required to represent design variables of RC beam. However, the number of design variables of RC beam is too many to represent by the action of conventional DQN. To solve this problem, multi-agent DQN was used in this study. For more effective reinforcement learning process, DDQN (Double Q-Learning) that is an advanced version of a conventional DQN was employed. The multi-agent of DDQN was trained for optimal structural design of RC beam to satisfy American Concrete Institute (318) without any hand-labeled dataset. Five agents of DDQN provides actions for beam with, beam depth, main rebar size, number of main rebar, and shear stirrup size, respectively. Five agents of DDQN were trained for 10,000 episodes and the performance of the multi-agent of DDQN was evaluated with 100 test design cases. This study shows that the multi-agent DDQN algorithm can provide successfully structural design results of RC beam.

Keywords : Multi-Agent reinforcement learning, Deep Q-Network, Optimal structural design, Reinforced concrete beam

1. 서론

근래에 들어와서 기계학습은 다양한 공학 분야에 적용되어 의미 있는 성과를 내고 있다. 특히 구조공학분야에 기계학습이 적용되어 연구된 사례는 최근 들어서 기하급수적으로 늘어나고 있다[1]. 구조공학분야에 기계학습이 적용된 1989년에서 2022년까지 발표된 485개의 논문을 분석한 결과 인공신경망을 이용한 연구가 가장 많이 수행되었고 그 외에 부스팅 알고리즘, 서포트 벡터머신, 랜덤 포레스트, 결정 트리 등의 다양한 기계학습 기법이 구조공학 분야에 적용되었다. 구조공학분야에 적용된 기계학습기법은 대부분 학습용 데이터셋을 사용해서 구조물이나 구조부재, 또는 재료의 특성 및 거동을 예측하는 평가모델 개발에 활용되어 왔고 이는 기계학습의 종류 중 지도학습에 해당된다[2,3]. 지도학습의 범주에 포함되는 기계학습 알고리즘을 사용해서 우수한 성과를 내기 위해서는 다양하고 풍부한 양질의 학습용 데이터셋이 필요하다. 구조공학 분야에서 사용되는 학습용 데이터셋은 구조 및 재료 실험이나 복잡한 비선형 해석 등을 통해서 구축되는 것이 일반 적이다. 그러나 이러한 방법으로 기계학습 모델을 훈련시키기 위한 양질의 데이터셋을 구축하기 위해서는 많은 노력과 비용이 요구된다.

기계학습의 한 분야인 강화학습은 지도학습의 핵심요소인 학습용 데이터셋을 구축할 필요 없이 환경과 에이전트간의 상호작용을 통해서 보상을 최대로 받을 수 있는 행동을 하도록 에이전트를 학습시킨다. 이러한 강화학습은 자율주행자동차, 로보틱스, 게임 등 제어공학 분야에서 널리 사용되고 좋은 성과를 내고 있다. 최근에는 구조제어공학 분야에서도 제어알고리즘 개발에 강화

* 종신회원, 영남대학교 건축학부 교수, 공학박사
School of Architecture, Yeungnam University
** 교신저자, 종신회원, 선문대학교 건축학부 교수, 공학박사
Division of Architecture, Sunmoon University.
Tel: 041-530-2315 Fax: 041-530-2839
E-mail : hskim72@sunmoon.ac.kr

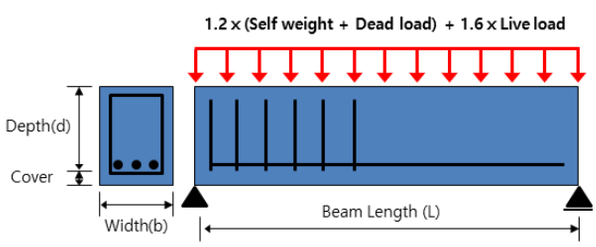
학습을 활용하여 우수한 성과를 내고 있다.

기계학습이 구조공학분야에서 활용되는 분야를 보면 구조부재의 거동과 구조재료의 특성을 예측하는데 주로 활용되었고 그 다음으로 구조물의 손상도 평가 및 모니터링에 활용되었다[1]. Thai(2022)가 분석한 결과에 의하면 구조해석 및 구조설계 분야에 기계학습이 적용된 연구는 약 11%에 불과하였다[1]. 이러한 연구는 대부분 최적의 해석 및 설계 결과를 이용해서 구성한 학습용 데이터셋을 이용하여 지도학습을 통해서 수행되었다. 또한 유전자알고리즘이나 다양한 최적화 기법을 이용한 구조부재 최적설계에 대한 연구도 수행되었다[4,5]. 그러나 현재까지 구조공학 분야에 기계학습 기법을 이용해서 수행된 다양한 연구를 검토해 보면 다른 주제에 비하여 구조물의 최적 설계에 대한 연구는 아직까지 상대적으로 매우 부족한 편이다. 특히 구조부재의 최적설계를 위하여 강화학습을 활용한 연구는 Jeong과 Jo(2021)의 연구가 거의 최초의 연구로 판단될 정도로 현재 초기 단계의 연구가 수행되고 있는 실정이다.

이러한 배경을 바탕으로 본 연구에서는 강화학습을 이용하여 구조부재의 최적설계 기술을 개발하였고 개발된 기술의 적용성 및 효율성을 검토하였다. 이를 위해서 설계대상 예제로 철근콘크리트(RC) 보를 선택하였다. 설계기준으로는 ACI(American Concrete Institute) 318을 적용하였다. 강화학습 알고리즘 중 DQN(Deep Q-Network)은 2015년 딥마인드가 하나의 알고리즘[7]으로 여러 개의 아타리 2600 게임을 사람보다 더 잘하는 수준으로 플레이하는 성과를 발표하면서 주목받기 시작하였고 다양한 문제에 적용되어 우수한 성과를 나타내었다. 그러나 본 연구에서 최적화 대상으로 선택한 RC 보의 설계변수의 선택범위가 넓기 때문에 일반적인 DQN을 본 연구에서 사용하기에는 부적절하다. 따라서 본 연구에서는 다중 에이전트 DDQN(Double Q-Learning)을 강화학습 알고리즘으로는 사용하였다[8]. 강화학습 환경은 RC 보의 재료비를 최소화시키면 큰 보상을 받을 수 있도록 설계하였다. 학습된 에이전트의 RC 보 구조설계 성능을 평가하기 위해서 100개의 설계조건에 대한 결과를 검토하였다. 본 연구에서 제안된 다중 에이전트 DDQN을 이용한 RC 보 최적설계기법의 효율성을 검증하기 위해서 선행 연구[6]의 연구결과와 비교·검토하였다.

2. RC 보 설계를 위한 강화학습 환경

본 연구에서는 RC 보를 구조설계의 대상 부재로 선택하고 이를 사용하여 강화학습을 수행할 환경을 구성하였다. 사용한 RC 보는 선행연구6)의 설계 대상과 동일하게 구성하여 본 연구에서 제안한 기법의 효용성을 비교·검토할 수 있도록 하였다. 설계 대상 RC 보의 구성을 〈Fig. 1〉에 나타내었다. 구조부재의 최적설계에 대한 연구에서는 일반적으로 보의 길이와 하중 등이 주어진 상태에서 최적 단면을 설계하는 것이 일반적이다. 그러나 본 연구에서는 그림에서 보는 바와 같이 보의 길이 및 하중이 임의의 값으로 변경할 때마다 그 조건에 최적화된 RC 보의 단면을 결정하도록 한다. 설계시 적용한 고정하중 및 활하중과 보 길이의 범위를 〈Table 1〉에 나타내었다.



〈Fig. 1〉 Configuration of example RC beam

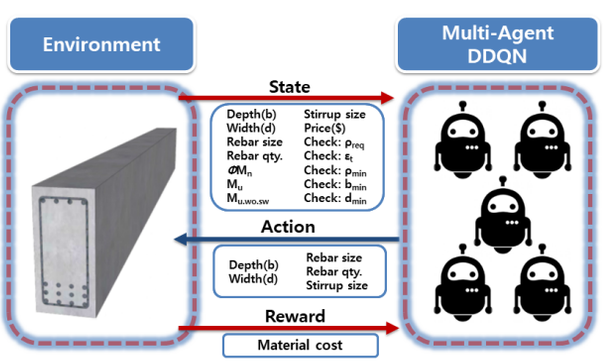
〈Table 1〉 Material properties and load values

Design Variable	Value	Unit
Dead load	0.5~3.0	kip/ft
Live load	0.5~3.0	kip/ft
Beam length	15~30	ft
Conc. comp. strength	4	ksi
Rebar strength	60	ksi
Rebar elastic modulus	29,000	ksi

콘크리트의 압축강도, 보강철근의 항복강도 및 탄성계수도 표에 함께 나타내었다. 본 연구에서는 ACI 318을 설계 기준으로 사용하였고 단위도 그에 맞추어 사용하였다.

RC 보 설계 강화학습 환경에서는 매 에피소드마다 〈Table 1〉에 나타낸 범위에서 임의의 하중과 보 길이를 갖는 예제를 생성하여 해당 조건에 대해 최적 설계를 수행하도록 한다. 설계대상이 단순보이므로 최대 소요 휨모멘트 및 전단력을 쉽게 계산할 수 있다. 이때 계산되

는 부재력은 아직 보의 단면이 결정되기 전이기 때문에 자중을 제외한 상태($M_{u,wo,sw}$)로 계산된다. 이후 보의 단면이 설계된 후 계산되는 실제 소요 휨모멘트(M_u) 및 전단력(V_u)에 대해서 설계가 수행된다. 본 연구에서 수행하는 RC 보 설계를 위한 강화학습 환경을 구성하는 상태(state), 행동(action), 보상(reward)을 〈Fig. 2〉에 나타내었다. 상태는 그림에서 보는 바와 같이 보의 춤(d), 너비(b), 보강철근 크기, 보강철근 개수, 설계강도, 소요강도, 자중 제외 소요강도, 스터럽 크기, 재료비, 균형철근비 검토, 인장변형률 검토, 최소철근비 검토, 최소 폭 검토, 최소 춤 검토의 14개로 구성된다. 이러한 상태가 다중 에이전트 DDQN 알고리즘에 전달되면 입력된 상태를 바탕으로 최적의 행동을 수행하는데 본 연구에서 선택된 행동은 RC 보 설계 결과이다. 즉, 보의 춤, 너비, 보강철근 크기, 개수 및 스터럽 크기가 에이전트의 행동으로 결정된다. RC 보의 최적설계는 ACI 318의 설계조건을 만족시키면서 최소한의 재료비를 사용하는 것을 목적으로 한다. 따라서 강화학습 환경은 재료비를 최소화할수록 더 큰 보상을 받도록 설계하였다.



〈Fig. 2〉 Configuration of RL environment

강화학습 환경의 상태에 따른 에이전트의 행동에 의해 결정되는 설계변수는 구조설계상 합리적인 범위의 값으로 선택되어야 한다. 이를 위해 본 연구에서는 에이전트가 선택할 수 있는 행동의 범위를 〈Table 2〉에 나타내었다. 보 단면의 춤과 너비는 각각 2 inch와 1 inch 간격으로 선택할 수 있도록 하였다. 주철근은 #6~#10, 스터럽은 #3~#5의 철근 중 선택할 수 있도록 하였고 주철근의 개수는 2~6개 사이에서 선택하도록 하였다.

〈Table 2〉 Section dimensions and rebar

Design Variable	Range	Unit
Section depth	10~60	in.
Section width	10~20	in.
Main rebar size	#6,7,8,9,10	-
Number of main rebar	2~6	EA
Stirrup size	#3,4,5	-

RC 보 강화학습 환경에서는 에이전트의 행동에 의해서 단면을 설계하고 설계된 단면의 성능을 바탕으로 보상을 계산해서 에이전트가 한 행동이 좋은 행동인지 그렇지 않은지를 학습을 시킨다. 보상을 계산할 때 강화학습 환경에서는 보의 설계결과를 두 단계로 나누어 평가한다. 첫 번째 단계는 주어진 설계기준을 만족시키고 있는지 판단한다. 대표적으로 설계된 RC 보의 설계강도(ϕMn)가 소요강도(Mu)보다 큰지 검토한다. 그 이외에도 ACI 318에서는 최소 철근비, 최소 보 폭, 순인장변형률, 보강근 최소단면적, 최대 보 폭-유효 깊이 비, 최소 보 춤의 6가지 설계기준을 만족해야 한다. 두 번째 단계는 설계기준을 만족시킨 RC 보의 재료비를 평가한다. 설계기준을 만족시켰다면 재료비가 작게 계산된 행동에 더 많은 보상을 준다. 첫 번째 단계는 반드시 만족시켜야할 기준이므로 이를 지키지 못할 때 음의 보상을 주도록 하였고 두 번째 단계는 경제성과 관련된 부분으로 재료비가 작으면 작을수록 양의 보상을 크게 하였다. 보상 계산을 위한 상세한 식은 선행연구6)에서 나타낸 내용을 사용하였다.

설계된 RC 보의 재료비는 아래의 식 (1)에 의해서 계산된다.

$$C_t = C_c + C_s \tag{1}$$

여기서 CC와 CS는 각각 콘크리트와 보강철근의 재료비를 의미한다. 콘크리트의 재료비는 다음과 같이 계산된다.

$$C_c = (bhL - V_r)P_c \tag{2}$$

여기서 b, h, L, V_r과 P_c는 각각 보의 폭, 깊이, 길이, 철근의 부피 그리고 단위 부피당 콘크리트의 가격을

의미한다. 철근의 재료비는 다음의 식으로 계산된다.

$$C_s = (A_s L + V_{stirr.}) P_s \quad (3)$$

여기서 A_s, L, P_s와 $V_{stirr.}$은 각각 주철근 단면적, 보 길이, 단위 부피당 철근의 가격 그리고 스터럽의 부피를 의미한다. 본 연구에서는 재료비를 계산하기 위해서 Lepš와 Šejnoha(2003)의 연구에서 사용한 값을 사용하였으며 콘크리트의 가격은 1.67 $/ft3 (59 $/$m^3$) 이고 철근의 가격은 490 $/$ft^3$(17,300 $/$m^3$)이다. 설계된 RC 보의 재료비에 의해서 계산되는 양의 보상은 임의로 결정되는 하중의 크기나 보의 길이에 따라서 크게 달라지므로 일정한 기준에 의한 정규화가 필요하다. 선행연구[6]에서는 다양한 하중 및 보의 길이 조건에 대해서 구조설계 전문가에 의해 설계된 RC 보를 사용해서 근사 최적비용(near-optimal cost)을 계산하고 이를 이용해서 강화학습의 보상을 정규화하였다. 강화학습 에이전트에 의해서 설계된 RC 보의 재료비가 근사 최적비용과 같으면 보상을 10으로 하였고 이보다 작으면 10보다 크게, 이보다 크면 10보다 작게 보상을 계산하였다. 즉, 강화학습 에이전트의 행동에 의해서 설계된 RC 보의 보상이 10보다 클수록 사용된 재료비가 전문가에 의해서 설계된 근사 최적 비용보다 작다는 것을 의미한다.

3. 다중 에이전트 강화학습의 개요

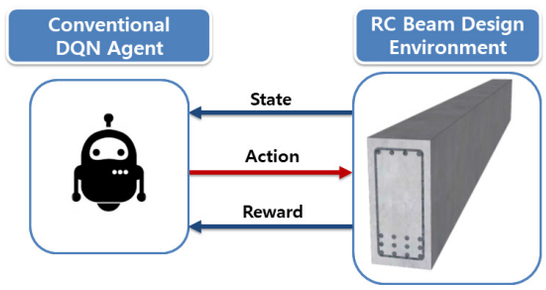
현재까지 다양한 강화학습 알고리즘이 개발되어 다양한 문제해결에 적용되고 있다. 대표적인 강화학습 알고리즘 에이전트의 타입 및 행동공간의 특성을 〈Table 3〉에 나타내었다. 초기에는 DQN을 비롯한 가치기반 강화학습 알고리즘이 많이 사용되다가 최근에는 정책기반 또는 가치네트워크와 정책 네트워크를 동시에 사용(actor-critic)하는 알고리즘이 많이 사용되고 있는데 대표적으로 DDPG가 여기에 포함된다. 강화학습 에이전트가 최적의 행동을 결정하게 되는 행동공간은 표에서 보는 바와 같이 크게 이산형(discrete) 공간과 연속형(continuous) 공간으로 나누어진다. 본 연구에서 강화학습 에이전트의 행동은 〈Table 2〉에 나타낸 값의 범위 내에서 선택을 해야 한다. 표에서 보는 바와 같이 주철근의 크기와 개수, 스터럽의 크기 등은 이산형 변수이다. 보의 폭과 깊이는 연속형 변수로 보일 수도 있지만 실제 건축적 제한조건이나 시공성 등을 고려하면 표현할 수 있는 길이의 값은 이산적일 수 밖에 없다. 따라서 본 연구의 대상인 RC 보 구조설계 환경에서는 이산형 변수를 행동으로 표현할 수 있는 알고리즘을 사용하는 것이 바람직하다. 〈Table 2〉에 나타낸 값의 범위 내에서 선택할 수 있는 행동의 개수는 보의 춤 26개(2 inch 간격), 보의 폭 11개(1 inch 간격), 주철근 크기 5개, 주철근 개수 5개, 스터럽 크기 5개로 총 50개이다. 이렇게 선택가능한 행동의 개수가 많으면 DQN과 같은 강화학습 알고리즘의 네트워크가 커지게 되고 최적의 행동을 찾는데 효율성이 떨어지게 된다. 따라서 동일한 대상 문제에 대해서 선행연구[6]에서는 연속형 변수를 출력하는 DDPG를 사용하였다.

〈Table 3〉 Agent types of RL algorithms

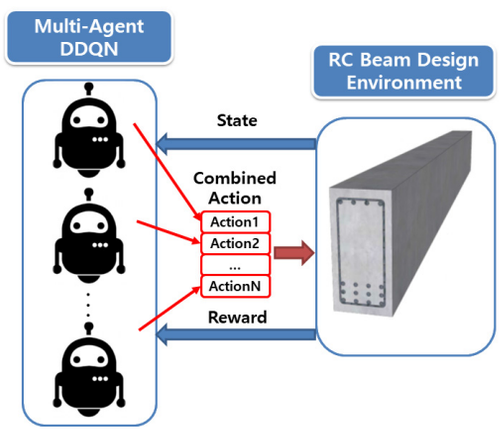
Agent	Type	Action Space
Q-Learning	Value-Based	Discrete
DQN	Value-Based	Discrete
SARSA	Value-Based	Discrete
PG	Policy-Based	Discrete, Continuous
AC	Actor-Critic	Discrete, Continuous
PPO	Actor-Critic	Discrete, Continuous
DDPG	Actor-Critic	Continuous
TD3	Actor-Critic	Continuous
SAC	Actor-Critic	Continuous

DDPG 행동 네트워크의 출력층에서는 tanh 활성화함수를 써서 출력의 범위를 −1과 1사이의 값으로 조정하였다. 앞서 설명한 바와 같이 DDPG의 행동은 연속형 실수로 출력되기 때문에 이를 이용해서 〈Table 2〉에 나타낸 범위 안의 설계변수 값을 선택하기 위해서는 적절한 후처리 작업이 필요하다. 또한 강화학습의 상태로 표현되는 설계변수 입력 값이 적절하게 정규화되지 않으면 DDPG의 출력은 −1과 1의 최소 또는 최대 값으로 집중되기 쉬운 특징을 가진다.

본 연구에서는 강화학습을 이용한 RC 보설계에 일반적인 DQN과 DDPG를 사용했을 때 발생하는 이러한 문제점을 해결하기 위해서 다중 에이전트 DDQN 알고리즘[8]을 사용하였다. 기존의 DQN 알고리즘과 다중 에이전트 DDQN 알고리즘의 구성을 〈Fig. 3〉에 나타내어 비교하였다.



(a) Conventional DQN



(b) Multi-agent DDQN

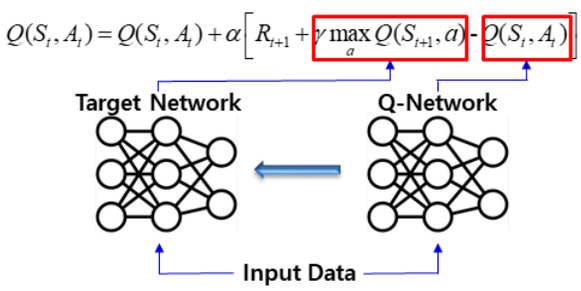
〈Fig. 3〉 Configuration of RL

다수의 에이전트가 협업 또는 경쟁하는 환경에서의 문제를 강화학습을 통해서 해결하려는 다중 에이전트 강화학습(Multi-Agent Reinforcement Learning, MARL)에 대한 연구가 다수 수행되어 왔다[10]. 다중 에이전트 강화학습을 효과적으로 활용하기 위해서는 에이전트간의 관계 모델링이나 에이전트간의 통신방법 및 신뢰할당 등 복잡한 모델링 과정을 통한 네트워크 설계가 필요하고 이는 쉽지 않은 일이다. 따라서 본 연구에서는 각각의 다중 에이전트가 서로 정보를 교환하지 않고 주어지는 동일한 상태에 대해서 각자가 담담한 설계변수만을 최적화시키는 단순한 방식의 다중 에이전트 강화학습 기법[9]을 사용하였다. 즉 〈Fig. 3(b)〉에 나타낸 바와 같이 각각의 다중 에이전트는 RC 보 강화학습 환경으로부터 동일한 상태를 입력받고 에이전트 당 하나의 설계변수만 행동으로 출력한 후 모든 에이전트의 행동을 묶어서 강화학습 환경에 전달한다. 이렇게 전달된 행동으로 설계된 보의 설계기준 만족여부 및 재료비를 평가하여 보상을 계산한 후 이를 각각의 다중 에이전트에게 전달하여 방금 전에 한 행동이 올바른 것인지 학습하게 한다. 본 연구에서는 〈Table 4〉에 나타낸 바와 같이 5개의 설계변수에 대응하는 5개의 다중 에이전트를 사용하였다. 이렇게 에이전트간의 정보교환 없이 각 에이전트가 담당하는 설계변수만 고려하여 최적 행동을 선택하면 되므로 네트워크 구성이 상대적으로 단순해질 수 있다. 선행연구[6]에서 사용한 DDPG는 64개의 노드를 가진 컨볼루션 신경망 레이어 8개를 사용하여 Actor와 Critic 네트워크를 구성하였다. 이에 비하여 본 연구에서 사용한 다중 에이전트는 모두 32개의 노드를 가진 일반 신경망 레이어 3개를 사용하여 네트워크를 구성하였다. 모든 다중 에이전트의 입력변수는 상태변수의 개수와 동일하게 14개로 하였고 출력변수는 〈Table 4〉에 나타낸 바와 같이 에이전트가 담당하는 설계변수에 따라 다르다.

〈Table 4〉 Actions of multi-agent

Multi-Agent	Actions	Design Variables
Agent1	10~60(step:2)	Section depth
Agent2	10~20(step:1)	Section width
Agent3	6,7,8,9,10	Main rebar size
Agent4	2,3,4,5,6	# of main rebar
Agent5	3,4,5	Stirrup size

DQN은 2015년에 발표된 이후에 DQN은 다양한 문제에 적용되어 우수한 성과를 나타내었다. 그러나 기존의 일반적인 Q-Learning 알고리즘을 기반으로한 DQN은 특정 조건에서 잘못된 행동의 가치를 과대평가해서 잘못된 방향으로 학습될 수 있는 문제점이 발견되었고 이러한 문제로 말미암아 몇몇의 적용 문제에서 좋은 성능을 보이지 못한 결과가 발표되었다. 이러한 문제를 해결하고자 Hasselt 등은 Double Q-Learning 알고리즘을 제안하여 잘못된 행동에 대한 과대평가를 줄여서 보다 다양한 문제에서 좋은 성능을 나타낼 수 있게 하였다[11]. DDQN(Double Q-Learning) 알고리즘의 기본 개념을 〈Fig. 4〉에 나타내었다.



〈Fig. 4〉 Concept of double DQN

일반적인 DQN에서 사용되는 max 연산자는 행동을 선택하고 평가하는데 같은 값을 사용한다. 이것은 과대평가된 값을 더욱 선택하게 하며, 너무 긍정적인 결과를 불러일으킨다. 이러한 문제를 해결하기 위해 DDQN 알고리즘에서는 선택과 평가 네트워크를 분리시켰다. DQN 알고리즘에서는 식(4)와 같은 방법으로 Q함수를 업데이트한다.

$$Y_t^{DQN} = R_{t+1} + \gamma \max Q(S_{t+1}, a|\theta_t) \quad (4)$$

여기서 R_{t+1}, S_{t+1}, γ는 행동 a에 대한 보상, 다음 스텝의 상태, 감가율을 의미한다. 이를 DDQN에서는 아래와 같이 선택과 평가를 분리해 표현한다. 여기서 θ_t 네트워크는 행동을 선택하고 $\overline{\theta_t}$는 정책을 평가한다. 학습시 경험을 통해 랜덤하게 하나의 네트워크만 학습을 진행하고 2개의 네트워크 가중치를 교환함으로써 대칭적으로 가중치 학습을 수행한다.

$$Y_t^{DDQN} = R_{t+1} + \gamma Q(S_{t+1}, \arg max\, Q(S_{t+1}, a|\theta_t)|\overline{\theta_t}) \quad (5)$$

본 연구에서는 이렇게 개선된 DDQN을 기반으로한 다중 에이전트를 구현하여 RC보 최적설계 알고리즘을 개발하였다.

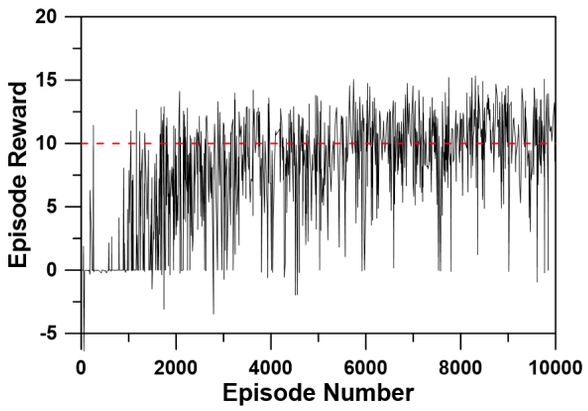
4. 다중 에이전트 강화학습을 이용한 RC보 최적 설계

다중 에이전트 DDQN 알고리즘을 이용한 RC 보 설계 강화학습 모델의 학습에 사용한 하이퍼파라미터를 〈Table 5〉에 나타내었다. 매 에피소드마다 5번의 스텝을 거쳐서 설계를 수행하였고 5번의 설계안 중 보상이 가장 큰 설계안을 선택하도록 하였다. 본 연구의 학습 대상인 RC 보 설계는 시간의존적인 문제가 아니다. 따라서 미래에 받게 될 보상에 대해서 가치를 줄여주는 감가율을 사용하지 않으므로 1의 값을 이용하였다.

〈Table 5〉 Hyperparameters for Multi-agent DDQN

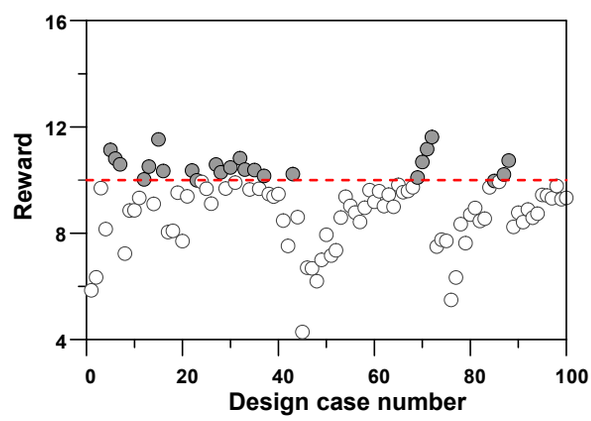
Item	Value
Number of agents	5
Learning rate	0.001
Target update frequency	1
Discount factor	1
Mini batch size	16
Activation function	Relu
Optimizer	Adam
Max. episode	10,000

학습 에피소드를 진행함에 따라서 멀티 에어전트 DDQN 알고리즘은 누적되는 보상이 많은 방향으로 발전하게 된다. 학습 에피소드의 증가에 따른 리워드의 변화이력을 〈Fig. 5〉에 나타내었다. 강화학습에서 각 단계에서 미래를 생각하지 않고 에이전트가 예측하는 최선의 행동만을 선택한다면 충분한 탐험(Exploration)이 되지 않아 국부 최적해에 빠질 위험이 있다. 이를 해결하기 위해 ϵ-greedy 정책을 사용하여 매 에피소드마다 생성되는 랜덤 값이 ϵ보다 작으면 임의의 행동을 선택하도록 하였다. ϵ의 최초 값은 1로 하였으며 감소율은 0.999로 하였고 0.005가 되면 더 이상 줄어들지 않게 하였다. 따라서 학습의 초기에는 큰 ϵ값에 의하여 탐험을 많이 해서 보상의 변동 폭이 크지만 에피소드가 증가할수록 ϵ값이 줄어들어서 현재의 지식을 활용(exploitation)하는 경향이 커지므로 변동폭이 줄어들게 된다. 〈Fig. 5〉에 나타낸 학습에 따른 보상의 변화이력을 보면 약 2,000 에피소드까지는 0의 보상이 매우 많이 나타나지만 그 이후로 평균적인 보상값이 점차 증가한다. 10,000에피소드 근처에서도 0 또는 음의 보상이 나타나는 것은 앞서 설명한 ϵ-greedy 정책에 의해서 선택된 임의의 행동 때문인 것으로 판단된다.

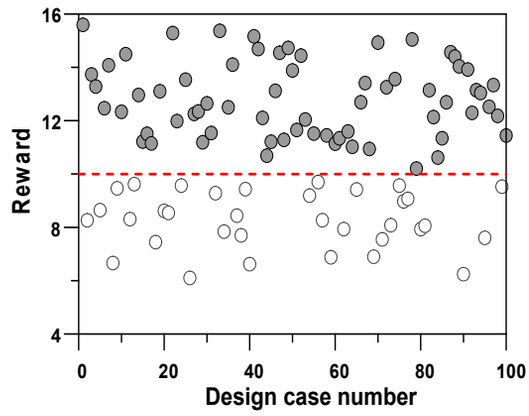


〈Fig. 5〉 Reward history along episode

학습된 다중 에이전트 DDQN 알고리즘의 성능을 검토해보기 위해서 임의로 생성된 100건의 설계 조건에 대해서 테스트 설계를 수행하였다. 본 연구에서 제안한 다중 에이전트 DDQN 알고리즘의 결과와 동일한 RC 보 강화학습 환경에서 학습된 DDPG 알고리즘의 결과[6]를 〈Fig. 6〉에 비교하였다. DDPG 알고리즘 경우에는 보상이 10이 넘는 경우가 22개이고 평균값은 9.04이다. 본 연구에서 적용한 다중 에이전트 DDQN 알고리즘에서는 평균이 11.27이고 보상이 10이 넘는 경우는 66 개였다. 이를 통해 본 연구에서 제안한 다중 에이전트 강화학습의 RC 보 설계의 적용성 및 학습 효율성이 우수한 것을 알 수 있다. 특히 선행연구에서는 100,000 에피소드의 학습을 진행한 것에 비해서 본 연구에서는 10%에 해당하는 10,000 에피소드만 학습하고도 더 우수한 설계결과를 나타내었다.



(a) Design results of single-agent DDPG[6]



(b) Design results of multi-agent DDQN
〈Fig. 6〉 Comparison of test design results

설계결과를 보다 구체적으로 확인하기 위해서 보의 길이가 15, 20, 25, 30 ft인 경우에 대해 〈Table 6〉에 나타내었다. 표에서 보는 바와 같이 보의 춤(d), 너비(b), 보강철근 크기, 개수 등이 본 연구에서 제안된 알고리즘에 의해서 적절하게 선택되는 것을 알 수 있다.

〈Table 6〉 Selected design results

Len	d	b	Rebar #	Rebar Qty.	Stirrup #	Cost($)	Reward
15	36	12	8	2	4	160.7	12.5
20	36	12	8	2	4	214.3	13.3
25	36	12	9	2	4	303.7	13.0
30	44	15	9	4	4	677.6	12.3

5. 결론

본 연구에서는 강화학습을 이용해서 RC 보의 최적설계 기술을 개발하였다. RC 보의 최적 설계변수를 선택하기 위해서는 이산형 강화학습 알고리즘을 적용하는 것이 적합하다. 이산형 강화학습 기법으로 유명한 일반적인 DQN 알고리즘은 많은 설계변수를 행동의 선택지로 학습을 하면 효율이 저하되기 때문에 본 연구에서는 각각의 설계변수를 담당하는 에이전트를 따로 학습시키는 다중 에이전트 DDQN 알고리즘을 사용하여 RC 보 설계를 학습시켰다. 개발된 모델의 활용성을 높이기 위해서 하중이나 보의 길이를 고정시키지 않고 학습 때마다 임의의 값을 사용한 설계조건에 대해 최적설계기법

을 학습하도록 하였다. 본 연구에서 제안한 기법의 우수성을 검증하기 위해서 동일한 강화학습 환경에서 DDPG 알고리즘을 이용해서 RC 보 설계 모델을 개발한 선행연구과 설계결과를 비교해 보았다. 그 결과 100,000 에피소드의 학습으로 얻은 DDPG 모델보다 10,000 에피소드의 학습으로 얻은 다중 에이전트 DDQN 모델이 구조 전문가가 설계한 결과를 기준으로 한 평가에서 3배 더 우수한 성능을 발휘하는 것을 확인할 수 있었다. 본 연구의 결과를 바탕으로 추후 RC 복근보 설계 및 기둥 등 다른 부재에 대한 설계모델 개발을 연속하여 수행할 계획이다.

감사의 글

본 논문은 2019년도 정부(과학기술정보통신부)의 재원으로 한국연구재단의 지원을 받아 수행된 연구임. (No. NRF-2019R1A2C1002385)

References

1. Thai, H.T., “Machine learning for structural engineering: A state-of-the-art review”, Structures, Vol.38, pp.448-491, 2022, doi: 10.1016/j.istruc.2022.02.003
2. Degtyarev, V.V. & Naser, M.Z., “Boosting machines for predicting shear strength of CFS channels with staggered web perforations”, Structures, Vol.34, pp.3391-3403, 2021, doi: 10.1016/j.istruc.2021.09.060
3. Feng, D.C., Wang, W.J., Mangalathu, S., Hu, G. & Wu, T., “Implementing ensemble learning methods to predict the shear strength of RC deep beams with/without web reinforcements”, Engineering Structures, Vol.235, 2021, doi: 10.1016/j.engstruct.2021.111979
4. Lee, C. & Ahn, J., “Flexural design of reinforced concrete frames by genetic algorithm”, Journal of Structural Engineering, Vol.129, pp.762-774, 2003, doi: 10.1061/(ASCE)0733-9445(2003)129:6(762)
5. Lee, K.S. & Geem, Z.W., “A new structural optimization method based on the harmony search algorithm”, Computers and Structures, Vol.82, pp.781-798, 2004, doi: 10.1016/j.compstruc.2004.01.002
6. Jeong, J.H. & Jo, H., “Deep reinforcement learning for automated design of reinforced concrete structures”, Computer-Aided Civil and Infrastructure Engineering, Vol.36, pp.1508-1529, 2021, doi: 10.1111/mice.12773
7. Volodymyr, M., Koray, K., David, S., Andrei, A.R., Joel, V., Marc, G.B., Alex, G., Martin, R., Andreas, K.F., Georg, O., Stig, P., Charles, B., Amir, S., Ioannis, A., Helen, K., Dharshan, K., Daan, W., Shane, L. & Demis, H., “Human-level control through deep reinforcement learning”, Nature, Vol. 518, pp. 529-533, 2015, doi: 10.1038/nature14236
8. Hafiz, A.M. & Bhat, G.M., "Deep q-network based multi-agent reinforcement learning with binary action agents", Computer Science, 2020, doi: 10.48550/arXiv.2008.04109
9. Lepš, M., & Šejnoha, M., “New approach to optimization of reinforced concrete beams”, Computers and Structures, Vol.81, pp.1957-1966, 2003, doi: 10.1016/S0045-7949(03)00215-3
10. Zhang, K., Yang, Z. & Başar, T., “Multi-agent reinforcement learning: a selective overview of theories and algorithms”, Studies in Systems, Decision and Control, Vol. 325, pp. 321-384, 2021, doi: 10.1007/978-3-030-60990-0_12
11. Hasselt, H.V., Guez, A. & Silver, D., “Deep reinforcement learning with double q-learning”, Proceedings of the Thirtieth AAAI Conference on Artificial Intelligence, pp.2094-2100, 2016, doi: 10.1609/aaai.v30i1.10295

■ Received : May 08, 2023
■ Revised : May 27, 2023
■ Accepted : May 27, 2023