

스마트 제어알고리즘 개발을 위한 강화학습 리워드 설계

Reward Design of Reinforcement Learning for Development of Smart Control Algorithm

김 현 수*
Kim, Hyun-Su

윤 기 용**
Yoon, Ki-Yong

Abstract

Recently, machine learning is widely used to solve optimization problems in various engineering fields. In this study, machine learning is applied to development of a control algorithm for a smart control device for reduction of seismic responses. For this purpose, Deep Q-network (DQN) out of reinforcement learning algorithms was employed to develop control algorithm. A single degree of freedom (SDOF) structure with a smart tuned mass damper (TMD) was used as an example structure. A smart TMD system was composed of MR (magnetorheological) damper instead of passive damper. Reward design of reinforcement learning mainly affects the control performance of the smart TMD. Various hyper-parameters were investigated to optimize the control performance of DQN-based control algorithm. Usually, decrease of the time step for numerical simulation is desirable to increase the accuracy of simulation results. However, the numerical simulation results presented that decrease of the time step for reward calculation might decrease the control performance of DQN-based control algorithm. Therefore, a proper time step for reward calculation should be selected in a DQN training process.

Keywords : Reinforcement learning, Smart TMD, Deep Q-network, Reward calculation, Seismic response reduction

1. 서론

지진하중에 대한 건축구조물의 동적응답을 저감시키기 위해서 동조질량감쇠기(Tuned Mass Damper; TMD)가 개발되어 널리 사용되고 있다¹⁾. 일반적으로 사용되고 있는 수동 동조질량감쇠기(Passive TMD; PTMD)는 제어 대상이 되는 건축구조물 모드형상의 고유진동주기에 맞추어 조율된다²⁾. 이 경우에 건축구조물의 동적특성이 변경되지 않는다면 PTMD는 지진응답을 매우 효율적으로 저감시킬 수 있다. 그러나 건축구조물의 사용상 각 실의 용도변경으로 인하여 탑재되는 질량이 변화되거나 건축물의 노후화 등에 의해서 강성이 변화됨으로 동적특성이 달라지면 PTMD의 제어성능이 급격하게 저하된다. 또한 PTMD는 조화하중에 대해서는 극적인 동적응답

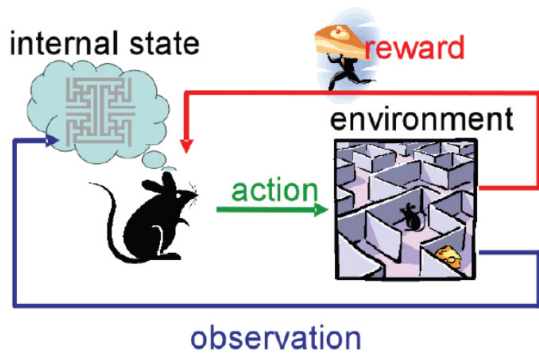
저감효과를 나타내지만 지진하중이나 풍하중과 같이 다양한 주파수 성분이 섞여있는 동적하중에 대해서는 동적응답 저감효과가 제한적이다. 이러한 PTMD의 단점을 보완하기 위해서 액추에이터와 같은 가력장치를 수동 감쇠기 대신에 사용한 능동 동조질량감쇠기(Active TMD; ATMD)가 개발되었다. ATMD는 능동제어장치를 포함하고 있으므로 엔지니어가 원하는 제어력을 어렵지 않게 제공할 수 있으므로 PTMD에 비하여 탁월한 동적응답 제어성능을 보였다³⁾. 그러나 이러한 능동제어장치를 운용하기 위해서는 많은 전력이 소모되므로 유지보수에도 상당한 비용이 소요되는 단점이 있다. 또한 적절한 제어력이 구조물에 가해지지 못하면 구조물의 응답을 오히려 증가시켜서 구조시스템을 불안정하게 할 가능성이 있다. 따라서 최근에는 수동제어 시스템이 갖고 있는 신뢰성(reliability)과 능동제어 기법의 장점인 적응성(adaptability)을 확보한 준능동 스마트 동조질량감쇠기(Smart TMD; STMD)에 대한 연구가 활발히 진행 중이다⁴⁻⁶⁾.

이러한 STMD의 제어성능은 STMD의 기계적 최적 설계 뿐만 아니라 제어알고리즘의 성능에 의해서 크게

* 종신회원, 선문대학교 건축학부 교수
Division of Architecture, Sunmoon University
** 교신저자, 선문대학교 건설시스템안전공학과 교수
Department of Civil Infrastructure Systems and Safety Engineering, Sunmoon University
Tel: 041-530-2326 Fax: 041-530-2926
E-mail : kyyoon@sunmoon.ac.kr

좌우된다. STMD를 제어하기 위해서 groundhook 및 skyhook과 같은 전통적인 on-off 방식의 준능동 제어 알고리즘⁴⁾, 능동제어알고리즘 기반의 준능동 제어알고리즘^{7,8)}, 퍼지논리제어기 및 인공신경망과 같은 소프트 컴퓨팅기법 기반의 제어알고리즘^{9,10)} 등 다양한 제어알고리즘이 개발되었고 각각의 특성 및 제어대상에 적합한 효과적인 제어성능을 보여주었다. 특히 소프트 컴퓨팅기법 기반의 제어알고리즘은 복잡하고 불확실하며 애매한 조건 아래서 활용되어야 하는 시스템에 매우 우수한 성과를 나타내었다. 알고리즘 고유의 견실성과 비선형성 및 불확실성을 쉽게 다룰 수 있는 능력 때문에 MR (magnetorheological) 감쇠기와 같은 스마트 제어장치로 구성된 제어시스템에 적용되어 우수한 성과를 나타내었다.

근래에는 기계학습에 기반한 기술이 다양한 공학문제 해결에 활용되어 우수한 성과를 나타내고 있다. 다양한 기계학습 기법 중에 강화학습은 학습을 하는 에이전트가 주어진 환경(environment)에 대해 관측(observation)할 수 있는 상태(state)를 기반으로 어떤 행동(action)을 취하면 환경으로부터 얻는 보상(reward)을 기반으로 학습을 이어가게 하는 알고리즘으로 이 과정을 <Fig. 1>에 나타내었다.



<Fig. 1> Reinforcement learning process¹¹⁾

이러한 강화학습 기본 개념에서 환경을 지진하중을 받는 스마트 제어시스템이 설치된 구조물로 두고 관측 및 상태를 지진하중을 받는 구조물과 스마트 제어시스템의 동적응답으로 두면 학습을 하는 에이전트를 제어 알고리즘으로 두어 강화학습을 적용할 수 있다. 이때 에이전트가 환경을 대상으로 취하는 행동은 제어알고리즘에서 스마트 TMD로 전달되는 제어명령이 되고 환경으

로부터 얻는 보상은 지진응답의 감소량이 될 것이다. 이렇게 강화학습을 적용하여 스마트 구조제어시스템의 제어알고리즘을 개발하려는 시도¹²⁾가 있었지만 관련분야 연구는 국내외적으로 아직 미미한 단계이다.

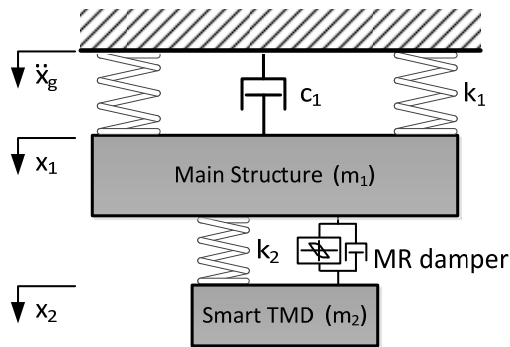
강화학습을 이용하여 스마트 구조제어시스템의 제어 알고리즘을 개발하는 과정에서 에이전트의 학습 성과에 영향을 미치는 다양한 하이퍼파라미터들이 존재한다. 특히 제어알고리즘의 역할을 하는 에이전트의 제어명령에 대해 환경이 제공하는 리워드를 어떻게 설계하느냐에 따라서 학습의 성과가 크게 달라질 수 있다. 스마트 구조제어시스템을 강화학습의 환경으로 하는 리워드 설계에 대한 국내외적으로 찾아보기가 쉽지 않으므로 본 연구에서는 리워드 설계에 영향을 미치는 주요 인자에 대해서 검토해보았다.

이를 위해서 스마트 TMD가 설치된 단자유도 구조물을 예제구조물로 선택하였다. 스마트 TMD를 구성하기 위해서 일반적인 PTMD에서 수동 감쇠기를 대신하여 MR 감쇠기를 사용하였다. 수치해석을 위한 지진하중으로는 강진지역 설계응답스펙트럼에 기반한 인공지진을 생성하여 사용하였다. 스마트 TMD가 설치된 단자유도 구조물에 대한 제어알고리즘을 강화학습을 통해서 개발하고자 할 때 일반적으로 유한요소 해석모델을 이용하여 강화학습의 환경을 구성한다. 인공지진하중을 받는 예제 구조물의 동적응답을 계산하기 위해서 상태공간 방정식을 구성하였고 수치적분을 위하여 Runge-Kutta 적분법을 사용하였다. 개발한 강화학습 기반의 스마트 제어알고리즘에서는 예제구조물의 지진피해를 최소화하기 위한 행동으로 스마트 TMD를 구성하는 MR 감쇠기로 전달될 최적의 명령전압을 생성한다. 이때 제어알고리즘이 최적의 성능을 나타내기 위해서 적절한 하이퍼파라미터의 선택 및 리워드 설계시 고려해야 할 사항에 대해서 검토하였다. 또한 개발된 스마트 제어알고리즘의 지진응답 저감성능을 수치해석을 통하여 검증하였다.

2. 예제구조물 및 수치해석 기법

본 연구에서는 강화학습을 이용한 스마트 제어알고리즘 개발을 위하여 예제구조물로서 대공간 돔구조물을 선택하였다. 선행연구를 통해서 대공간 돔구조물에 지진하중이 가해질 경우에 대부분의 동적응답은 구조물의

상하방향으로 발생한 다는 것을 확인할 수 있었다¹³⁾. 예제구조물의 지진응답을 계산하기 위한 시간이력해석시간을 줄이기 위해서 대공간 돔구조물의 동적특성을 모사할 수 있는 단자유도 해석모델을 사용하였고 지진하중이 지반을 거쳐서 돔구조물의 최상 중앙부에 유발하는 상하 동적응답을 표현하기 위해서 <Fig. 2>에 나타난 바와 같이 예제구조물을 모형화하였다.



<Fig. 2> Example structure with STMD

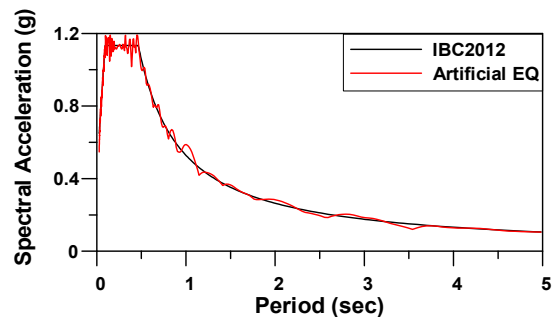
그림에 나타난 예제구조물에서 모사하고자하는 돔구조물은 고유진동수 3Hz를 갖는 것으로 가정하였다. 이를 구현하기 위해서 주구조물의 질량(m_1)과 강성을 각각 750kg 및 각 266,670N/m으로 하였고 감쇠비는 1.0%로 설정하였다. 스마트 TMD를 구성하는 질량(m_2)은 100kg으로 두었다. 스마트 TMD의 최적 진동수비는 동일 질량을 가지는 PTMD의 최적값을 사용하였고 이를 식(1)에 나타내었다¹⁴⁾. 이 식에서 μ 는 주구조물에 대한 TMD의 질량비(m_1/m_2)를 의미한다.

$$\frac{\sqrt{1-\mu/2}}{1+\mu} \tag{1}$$

식(1)에 의해서 계산되는 최적 진동수비를 사용하면 스마트 TMD를 구성하는 스프링 강성(k_2)를 결정할 수 있다. 전술한 바와 같이 본 연구에서 사용하는 스마트 TMD는 일반적인 PTMD를 구성하는 점성감쇠기를 대신하여 MR 감쇠기를 사용하여 구성된다. 이어서 소개할 인공지진하중을 사용하여 예제구조물에 최적의 성능을 나타내는 스마트 TMD를 구성하기 위한 MR 감쇠기 용량을 검토한 결과 1,000N으로 선택하였다. 스마트 TMD의 감쇠력을 실시간으로 제어하기 위해서 MR 감쇠기로

전달되는 전압의 범위는 0V~10V로 설정하였다. MR 감쇠기의 비선형 동적거동을 표현하기 위한 수치해석모델로는 <Fig. 2>에 나타난 바와 같이 비교적 단순하면서도 정확도가 높은 Bouc-Wen 모델¹⁵⁾을 사용하였다.

본 연구에서는 강화학습의 환경을 구성하기 위해서 강진지역의 특성을 나타내는 인공 지진하중을 작성하기 위하여 IBC2012 설계기준에 근거한 설계응답스펙트럼을 사용하였다. IBC2012에 의한 설계응답스펙트럼 작성 시 건물이 건설될 위치는 미국 Los Angeles로 선택하였고 이에 따라 S_s 는 1.70, S_1 은 0.61으로 결정되었다. Site class는 C로 선택하여 F_a 와 F_v 는 각각 1.0과 1.3으로 하였다. 이에 따라서 S_{MS} 는 1.70, S_{M1} 은 0.793, S_{DS} 는 1.1333, S_{D1} 은 0.5287로 계산되고 T_0 는 0.0933초 T_S 는 0.4665초로 각각 계산되었으며 T_L 은 30층 이상의 건물에서만 의미가 있으므로 본 연구에서는 고려하지 않았다. 이렇게 생성한 IBC2012에 근거한 설계응답스펙트럼을 <Fig. 3>에 나타내었다.

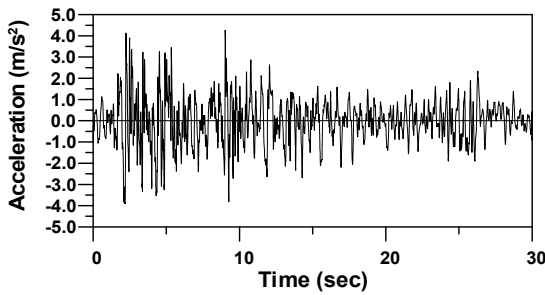


<Fig. 3> Response spectrum of artificial EQ

그림에서 보는 바와 같이 IBC2012로 생성된 설계응답스펙트럼의 단주기 영역 스펙트럼가속도는 1.13g로서 강진지역의 특성을 나타내는 것을 알 수 있다.

본 연구에서는 지각내부의 특성 및 지반 종류의 특성이 반영되어 기록된 역사지진을 가정된 초기의 지진파형으로 사용하여 인공지진을 생성하였다. 이를 위해서 가정된 초기 지진파형으로 사용되는 역사지진의 지진동파형과 동일한 위상특성을 가지면서 설계용 가속도 응답스펙트럼에 매우 근접한 응답스펙트럼을 갖는 인공지진을 작성하는 방법을 사용하였다. 이 방법은 초기지진의 개략적인 파형을 그대로 유지하면서 위상특성의 변화가 없고 설계용 가속도 스펙트럼에 대응하는 여러 종

류의 역사지진을 바탕으로 하는 인공지진을 손쉽게 작성할 수 있다. 인공지진 작성을 위해서 일반적으로 구조물의 지진응답 거동을 분석할 때 널리 사용되고 있는 El Centro (1940, NS) 지진하중을 사용하였다. 이 El Centro 지진하중의 개략적인 파형을 그대로 유지하면서 위상특성의 변화가 전혀 없고 설계용 가속도 스펙트럼에 대응하는 인공지진하중을 생성한다. 이러한 방법으로 작성한 인공지진동의 가속도 시간이력을 <Fig. 4>에 나타내었고 응답스펙트럼을 생성하여 인공지진하중을 작성할 때 사용한 IBC2012에 근거한 설계응답스펙트럼과 함께 <Fig. 3>에 나타내었다. 그림에서 보는 바와 같이 인공지진하중의 가속도 응답스펙트럼과 설계응답스펙트럼이 전반적으로 잘 일치하는 것을 알 수 있다. 또한 <Fig. 4>에 나타낸 인공지진하중의 가속도 파형은 원래 El Centro 지진하중과 비슷하게 나타났다.



<Fig. 4> El Centro EQ based artificial EQ

본 논문에서 강화학습의 환경을 구성하기 위해 사용한 <Fig. 2>에 나타낸 예제구조물을 동적평형방정식으로 표현하면 다음과 같다.

$$M\ddot{x} + C\dot{x} + Kx = Df_c + Ef_e \quad (4)$$

여기서 $x = [x_1 \ x_2]^T$ 는 스마트 TMD와 주구조물의 변위를 나타내고 f_e 는 지진하중을 의미하며, M, C, K는 각각 예제구조물의 질량, 감쇠 및 강성행렬을 의미한다. 또한 D와 E는 스마트 TMD의 제어력과 지진하중의 작용위치를 나타내는 행렬이고, f_c 는 스마트 TMD의 제어력을 의미한다. 이러한 동적평형 방정식은 다음과 같은 상태공간방정식으로 나타낼 수 있다.

$$\dot{z} = Az + Bf_c + Hf_e \quad (4)$$

여기서, $A = \begin{bmatrix} 0 & I \\ -M^{-1}K & -M^{-1}C \end{bmatrix}$ 는 시스템행렬, $B = \begin{bmatrix} 0 \\ M^{-1}D \end{bmatrix}$ 와 $H = \begin{bmatrix} 0 \\ M^{-1}E \end{bmatrix}$ 는 위치행렬이고, $z = [x_1 \ x_2 \ \dot{x}_1 \ \dot{x}_2]^T$ 는 상태벡터이다. 이러한 상태공간방정식을 사용하여 시간이력해석을 수행하기 위해서는 수치적분이 필요하다. 본 연구에서는 다양한 수치적분법 중에서 널리 사용되고 있는 Runge-Kutta 기법을 사용하였다. 독일의 수학자 카를 다비트 룬게와 마르틴 빌헬름 쿠타가 개발하였고 일반적으로 식(5)에 나타낸 바와 같이 4차항까지 구하여 사용하는 방법이 많이 쓴다.

$$y_{i+1} = y_i + (a_1k_1 + a_2k_2 + a_3k_3 + a_4k_4)h \quad (5)$$

여기서, k_n 은 다음과 같이 정의된다.

$$k_1 = f(t_i, y_i)$$

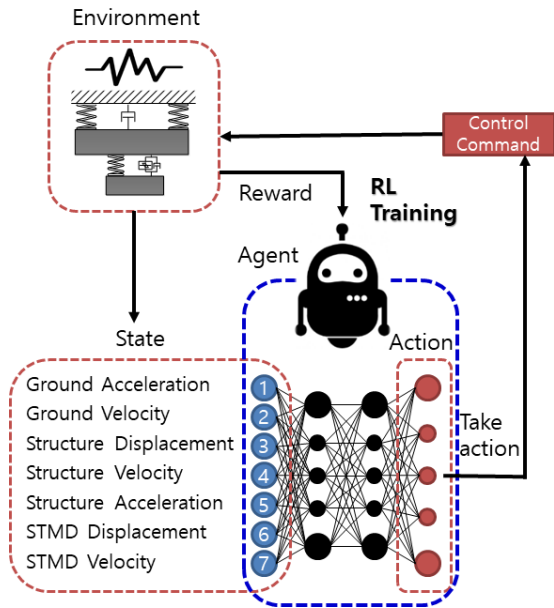
$$k_2 = f(t_i + \frac{1}{2}h, y_i + \frac{1}{2}k_1h)$$

$$k_3 = f(t_i + \frac{1}{2}h, y_i + \frac{1}{2}k_2h)$$

$$k_4 = f(t_i + h, y_i + k_3h)$$

3. 강화학습 구성 및 리워드 설계

여러 강화학습 알고리즘 중에서 본 연구에서는 Deep Q-Network (DQN)¹⁶⁾을 사용하였고 지금까지 다양한 공학적 문제해결에서 우수한 성능을 보여주었다. 강화학습을 이용하여 한 스마트 TMD를 제어하는 알고리즘을 개발하기 위해서는 앞 절에서 구현한 강화학습 환경에 대해서 적절한 관측 및 상태 값을 선택하고 이를 바탕으로 리워드 계산 및 에이전트를 학습시키는 과정이 필요하다. 이와 같은 본 논문에서 적용한 강화학습의 전반적인 구성에 대해서 <Fig. 5>에 나타내었다.



〈Fig. 5〉 Configuration of reinforcement learning for smart TMD

그림에 나타난 바와 같이 지진하중을 받는 스마트 TMD가 설치된 예제구조물로 구성된 환경에서 7개의 동적응답(①지반가속도, ②지반속도, ③주구조물 변위, ④주구조물 속도, ⑤주구조물 가속도, ⑥STMD 변위, ⑦STMD 속도)을 에이전트가 관측하는 상태로 선택하였다. 본 연구에서는 사용하는 인공지능망으로 구성된 DQN 에이전트는 7개의 동적응답 상태를 입력으로 하고 스마트 TMD로 전달되는 명령전압을 행동으로 출력한다. 이때 MR 감쇠기의 입력전압 범위는 0~10V이므로 이 사이를 1V간격으로 나누어 총 11개 간격의 명령전압 1개를 DQN 에이전트가 출력하는 행동으로 하여 매 스텝마다 최적의 행동을 선택하도록 학습한다. 이렇게 최적의 행동을 선택하도록 학습하는 과정에서 선택한 행동에 대한 리워드를 어떻게 설계하느냐가 학습의 효율성을 크게 좌우하게 된다.

본 연구에서는 구조물에 발생할 수 있는 피해와 가장 직접적으로 연관이 있다고 판단되는 최대 변위응답을 이용해서 리워드를 설계하였고 이를 식(6)에 나타내었다. 일반적으로 딥러닝에서는 학습의 효율성을 높이고 지역 최적값에 빠지는 가능성을 줄이기 위해서 데이터를 정규화(Normalization)하여 사용한다. 강화학습에서는 리워드를 최대화하는 방향으로 에이전트를 학습시키므로 다음 식과같이 리워드의 분모에 구조물의 응답을

두어 응답이 저감되면 리워드가 커지게 하였다. 주구조물의 최대응답이 커질 때 더 많은 페널티를 주기 위해서 응답을 제곱하였고 정규화된 주구조물의 최대응답은 0~1 사이에 있으므로 제곱하면 값이 더 줄어들므로 10을 곱하였다.

$$R = 1/(\text{Normalized Peak Disp.} \times 10)^2 \quad (6)$$

본 연구의 강화학습 환경에서는 0.001초 간격으로 수치해석을 수행하였다. 따라서 0.001초 간격으로 환경의 상태가 변하는데 이렇게 변하는 상태를 바탕으로 리워드를 계산하고 에이전트의 행동을 결정한다는 것은 스마트 TMD를 0.001초 간격으로 제어한다는 것과 같은 의미이다. 그러나 실제로 스마트 TMD를 제어하기 위해서는 가속도계와 같은 계측장치로 얻은 응답을 입력으로 하여 강화학습 에이전트가 인공지능망 계산 후 출력한 명령전압을 MR 감쇠기로 보내 제어가 이루어진다. 이때 시간간격을 0.001초와 같이 매우 짧은 시간으로 설정한다면 시간지연(time lag)이 발생하여 원하는 제어를 할 수 없는 상황이 발생할 수 있다. 또한 제어하는 시간간격이 구조물의 고유진동주기에 비해 너무 짧다면 구조물의 동적거동 상태를 적절하게 판단하지 못할 수가 있다. 따라서 본 연구에서는 해석은 0.001초의 시간간격으로 수행하였지만 리워드 계산을 위한 시간간격을 변화시켜서 학습을 시킨 후 이에 따른 주구조물의 최대 변위응답을 비교하여 보았고 이를 〈Table 1〉에 나타내었다. 이때 리워드 계산을 하는 시간간격 동안에는 에이전트의 행동(명령전압)은 일정하게 하였다. 그리고 학습에 사용한 하이퍼파라미터는 〈Table 2〉에 나타난 바와 같다.

〈Table 1〉 Training results based on time steps for reward calculation

Time step	0.001sec	0.01sec	0.1sec	1sec
Peak disp.	3.13cm	2.94cm	2.99cm	2.66cm

〈Table 2〉 Hyperparameter for training

Item	Value
Learning rate	0.001
Target update frequency	5
Discount factor	0.99
Mini batch size	256
Activation function	Relu
Optimizer	Adam
Gradient threshold	1
Max. episode	10,000

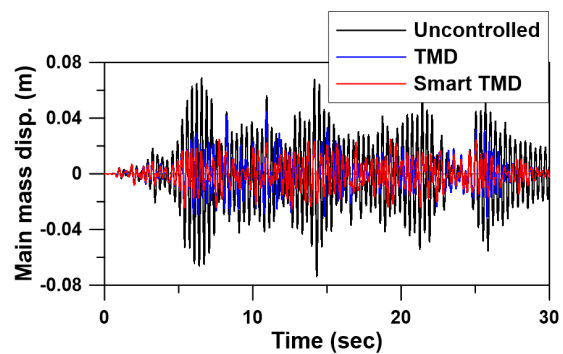
〈Table 1〉을 보면 리워드를 계산하는 시간 간격에 따라서 제어성능이 크게 달라지는 것을 알 수 있다. 해석의 정확도를 위해서는 시간 간격을 줄일수록 좋지만 제어알고리즘을 개발하는 강화학습 리워드 계산을 위해서는 오히려 시간 간격을 증가시켜 적절한 값을 찾는 것이 필요하다. 본 연구에서 사용한 예제에서는 1초 간격으로 리워드를 계산하고 제어명령을 결정하는 것이 가장 우수한 결과를 나타내는 것으로 나타났다. 〈Table 2〉에 나타난 하이퍼파라미터 이외에도 강화학습 에이전트를 구성하는 인공신경망의 노드 수와 레이어 수의 최적값을 선정하는 것이 필요하다. 이를 위해 강화학습 에이전트의 인공신경망 노드 수와 레이어 수에 따른 주구조물 최대응답 변화를 〈Table 3〉에 비교하였다. 일반적으로 인공신경망이 깊거나(deep) 넓으면(wide) 더 복잡한 판단을 효과적으로 할 수 있지만 과도하게 큰 인공신경망은 학습에 비효율적이거나 과적합 현상이 나타날 수 있다. 표를 보면 노드 수와 레이어 수를 각각 20개와 2개일 때 가장 효과적인 제어를 하는 것을 확인할 수 있으므로 본 연구에서는 이 값들을 사용하였다. 선택된 나머지 파라미터들은 시행착오법으로 결정하였다.

〈Table 3〉 Design parameters of agent

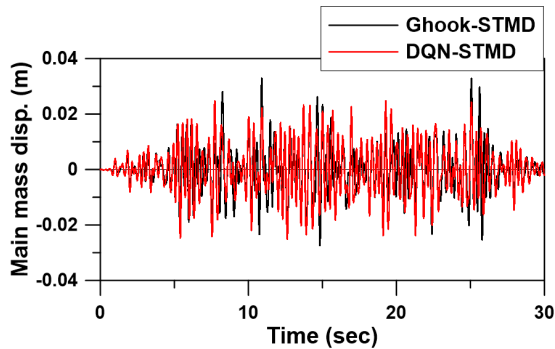
Number of node	Number of hidden layer	Peak displacement
20	1	3.14
20	2	2.63
30	2	2.78
40	2	2.70
40	3	2.76

4. 스마트 제어알고리즘의 성능 검토

앞서 설명한 방법대로 예제구조물에 설치된 스마트 TMD를 제어하기 위해서 10,000에피소드의 DQN 강화 학습을 수행하여 제어알고리즘을 개발하였고 개발한 알고리즘의 지진응답 제어성능을 검토하였다. 제어성능의 검토는 지진하중에 대한 예제 구조물의 피해를 가장 잘 표현할 수 있는 주 구조물의 변위응답을 사용하였다. 개발한 제어알고리즘에 의해서 제어되는 스마트 TMD의 제어성능 비교를 위하여 제어하지 않은 경우와 최적의 수동 TMD를 사용한 경우의 시간이력변위를 〈Fig. 6〉에 함께 나타내었다. 또한 개발된 DQN 기반 스마트 제어알고리즘의 제어성능을 비교하기 위해서 비교적 간단한 계산을 통해서 효과적인 제어성능 안정적으로 발휘하여 널리 사용되고 있는 준능동 제어알고리즘인 그라운드훅(ground-hook) 제어알고리즘을 사용한 결과와 함께 〈Fig. 7〉에 나타내었다. 〈Fig. 6〉을 보면 알 수 있듯이 비제어시보다 수동 TMD를 사용했을 때 응답이 크게 줄어들었고 스마트 TMD를 사용하면 더 크게 줄일 수 있었다. DQN 기반 제어알고리즘을 사용하면 그라운드훅 제어알고리즘을 사용했을 때 약 11초 부근과 26초 부근에서 크게 튀는 응답을 효과적으로 제어할 수 있었고 전반적으로 우수한 제어성능을 얻을 수 있었다.



〈Fig. 6〉 Comparison of main displacement



〈Fig. 7〉 Comparison of control algorithm

위의 그림을 명확히 파악하기 어려운 최대 응답치를 정량적으로 비교하기 위하여 제어케이스별 4개 응답의 최대치와 각 응답의 비를 계산하여 〈Table 4〉에 나타내었다.

〈Table 4〉 Comparison of peak responses

Case	Peak disp.	Ratio
Uncontrolled(U)	7.38cm	
PTMD(P)	4.45cm	0.60 (P/U)
STMD-Ghook(G)	3.29cm	0.74 (G/P)
STMD-DQN(D)	2.52cm	0.76 (D/G)

응답의 비를 보면 제어하지 않았을 때의 응답을 PTMD를 사용하여 40% 더 줄일 수 있었고 PTMD응답은 그라운드혹으로 제어한 STMD를 이용하여 약 26% 더 줄일 수 있었다. 그리고 본 연구에서 개발한 DQN 기반 제어알고리즘은 그라운드혹 제어알고리즘에 비해서 약 24% 응답을 더 줄일 수 있음을 확인할 수 있었다.

5. 결론

본 연구에서는 스마트 TMD의 제어 알고리즘을 개발하기 위해서 강화학습 알고리즘인 DQN을 사용하였고 지진응답 제어성능을 검토하였다. 이를 위해 스마트 TMD가 설치된 단자유도 구조물을 예제구조물로 선택하고 인공지진하중을 가하여 수치해석을 수행하였다. 강화학습을 이용한 제어알고리즘의 개발과정에서 환경구축 및 상태선택과 함께 리워드 설계가 매우 중요한 역할을

하게 된다. 본 연구에서는 특히 리워드를 계산할 때 사용하는 시간간격을 수치해석할 때 사용하는 시간간격과 다르게 하여 그에 따르는 학습의 성과를 검토하였고 리워드 설계시 시간간격이 매우 중요한 요소임을 확인하였다. 본 연구를 통해서 개발한 스마트 TMD는 최적의 수동 TMD뿐만 아니라 기존에 널리 사용되는 준능동 제어알고리즘인 그라운드혹 제어알고리즘보다 매우 우수한 제어성능을 나타낸다는 것을 확인할 수 있었다. 본 연구에서 제안한 기법을 추후 보다 다양한 예제구조물 및 지진하중을 사용한 검토가 추후 필요하다고 판단된다. 또한 성능검토에서 시간이력변위와 최대 응답치를 결과로 분석하였는데 추후 실제 구조물의 보다 다양한 응답의 분석이 필요할 것이다.

감사의 글

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

References

1. Matta, E., "Performance of tuned mass dampers against near-field earthquakes", *Structural Engineering and Mechanics*, Vol.39, No.5, pp. 621-642, 2011, doi: <https://doi.org/10.12989/sem.2011.39.5.621>
2. Warburton, G.B., "Optimum absorber parameters for various combinations of response and excitation parameters", *Earthquake Engrg. and Struct. Dyn.*, Vol.10, pp.381-401, 1982., doi: 10.1002/eqe.4290100304
3. Huu, T.P., Miura, N. and Iba, D., "Multi active tuned mass dampers for earthquake-induced vibration response control of high rise building", *Journal of Mechanical Science and Technology*, Vol.18, 2022, doi: 10.1007/s12206-022-0304-6
4. Koo, J.H., Using magneto-rheological dampers in semiactive tuned vibration absorbers to

- control structural vibrations, Ph.D. Dissertation, Virginia Polytechnic Institute and State University, USA, 2003.
5. Kim, H.S. and Kang, J.W., "Seismic response control of retractable-roof spatial structure using smart TMD", *Journal of the Korean Association for Spacial Structures*, Vol.16, No.4, pp.91-100, 2016, doi: 10.9712/KASS.2016.16.4.091
 6. Bathaei, A., Zahrai, S.M. and Ramezani, M., "Semi-active seismic control of an 11-DOF building model with TMD+MR damper using type-1 and -2 fuzzy algorithms", *Journal of Vibration and Control*, Vol. 24, No. 13, pp. 2938-2953, 2018, doi: 10.1177/1077546317696369
 7. Yi, F., Dyke, S.J., Caicedo, J.M., and Carlson, J.D., "Experimental verification of multi-input seismic control strategies for smart dampers," *Journal of Engineering Mechanics*, ASCE, Vol. 127, No. 11, pp. 1152-1164, 2001, doi: 10.1061/(ASCE)0733-9399(2001)127:11(1152)
 8. Dyke, S.J., Spencer, B.F., Sain, M.K. and Carlson, J.D., "Modeling and control of magnetorheological dampers for seismic response reduction", *Smart Materials and Structures*, Vol. 5, pp. 565-575, 1996.
 9. Kim, H.S. and Kang, J.W., "Vibration control performance evaluation of hybrid mid-story isolation system for a tall building", *Journal of the Korean Association for Spacial Structures*, Vol.18, No.3, pp.37-44, 2018, doi: 10.9712/KASS.2018.18.3.37
 10. Kim, H.S. and Kang, J.W., "Seismic response control of retractable-roof spatial structure using smart TMD", *Journal of the Korean Association for Spacial Structures*, Vol.16, No.4, pp.91-100, 2016, doi: 10.9712/KASS.2016.16.4.091
 11. Vorman, M.C., Maximum likelihood inverse reinforcement learning, Ph.D. Dissertation, , The State University of New Jersey, USA, 2014, doi: 10.7282/T3GQ70C8
 12. Kim, H.S. and Kang, J.W., "Development of semi-active control algorithm using deep q-network", *Journal of the Korean Association for Spacial Structures*, Vol.21, No.1, pp.79-86, 2021, doi: 10.9712/KASS.2021.21.1.79
 13. Kim, H.S. and Kang, J.W., "Seismic Response Control of Spacial Arch Structures using Multiple Smart TMD", *Journal of the Korean Association for Spacial Structures*, Vol.16, No.1, pp.43-51, 2016, doi: 10.9712/KASS.2016.16.1.043
 14. Warburton, G.B., "Optimum absorber parameters for various combinations of response and excitation parameters", *Earthquake Engineering and Structural Dynamics*, Vol.10, pp.381-401, 1982, doi: 10.1002/eqe.4290100304
 15. Sues, R. H., Mau, S. T. and Wen, Y. K., "System identification of degrading hysteretic restoring forces", *Journal of Engineering Mechanics*, ASCE, Vol.114, No.5, pp.833-846, 1988, doi: 10.1061/(ASCE)0733-9399(1988)114:5(833)
 16. 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. and Demis, H., "Human-level control through deep reinforcement learning", *Nature*, Vol.518, pp.529-533, 2015, doi: 10.1038/nature14236
- Received : May 12, 2022
 - Revised : June 9, 2022
 - Accepted : June 10, 2022