(19) 대한민국특허청(KR)
(12) 등록특허공보(B1)

(45) 공고일자 2023년05월09일
(11) 등록번호 10-2531181
(24) 등록일자 2023년05월04일

(51) 국제특허분류(Int. Cl.)
G06N 20/00 (2019.01)
(52) CPC특허분류
G06N 20/00 (2021.08)
(21) 출원번호 10-2021-0027610
(22) 출원일자 2021년03월02일
심사청구일자 2021년03월02일
(65) 공개번호 10-2022-0123975
(43) 공개일자 2022년09월13일
(56) 선행기술조사문헌
Sindhu Padakandla et al. Reinforcement learning algorithm for non-stationary environments. Applied Intelligence (2020) 50:3590-3606, 2020.06.18.*
Sindhu Padakandla et al. A Survey of Reinforcement Learning Algorithms for Dynamically Varying Environments. arXiv:2005.10619v1 [cs.LG], 2020.05.19.
US20200043359 A1
*는 심사관에 의하여 인용된 문헌

(73) 특허권자
연세대학교 산학협력단
서울특별시 서대문구 연세로 50 (신촌동, 연세대학교)
(72) 발명자
김시호
인천광역시 연수구 송도과학로 85, 연세대학교 국제캠퍼스 진리관 C동(송도동)
(74) 대리인
민영준

전체 청구항 수 : 총 8 항 심사관 : 박성수

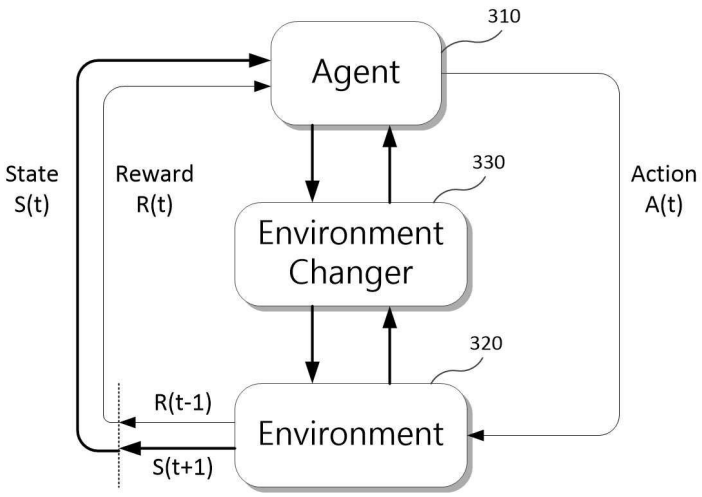
(54) 발명의 명칭 **인공지능 학습 장치 및 방법**

(57) 요 약

본 발명은 상태 데이터와 보상 데이터가 인가되면, 기설정된 액터 데이터에 따라 액션 데이터를 출력하는 에이전트 모듈, 에이전트 모듈로부터 액션 데이터가 인가되면, 기설정된 환경 데이터에 따라 인가된 액션 데이터에 대응하여 상태 데이터와 보상 데이터를 업데이트 하는 환경 모듈 및 에이전트 모듈 및 환경 모듈에 미리 설정되거

(뒷면에 계속)

대 표 도 - 도3



나 업데이트 되는 데이터 중 적어도 하나를 인가받아 기지정된 화면으로 구성하여 출력하고, 사용자에 의해 에이전트 모듈 및 환경 모듈에 미리 설정되거나 업데이트되는 데이터 중 적어도 하나의 데이터를 변경하기 위한 변경 데이터가 설정되어 인가되면, 인가된 변경 데이터에 대응하는 데이터를 변경 데이터로 대체하는 환경 변경 모듈을 포함하여, 강화 학습 중에 액터, 액션, 보상 및 상태 등을 다양하게 변화시킬 수 있을 뿐만 아니라, 다른 차원의 값으로 변화시킬 수 있도록 하여, 각종 예기치 못한 환경 변화에도 유연하게 대응할 수 있도록 학습시킬 수 있는 인공지능 학습 장치 및 방법을 제공할 수 있다.

이 발명을 지원한 국가연구개발사업

과제고유번호	1711117053
과제번호	2020-0-00056-001
부처명	과학기술정보통신부
과제관리(전문)기관명	정보통신기획평가원
연구사업명	정보통신방송연구개발사업
연구과제명	현실 세계에서 변화하는 상황에 따라 지속적으로 자가 개선하는 인공지능 기술 개발
기 여 율	1/1
과제수행기관명	연세대학교 산학협력단
연구기간	2020.04.01 ~ 2020.12.31

명 세 서

청구범위

청구항 1

상태 데이터와 보상 데이터가 인가되면, 기설정된 액터 데이터에 따라 액션 데이터를 출력하는 에이전트 모듈;

상기 에이전트 모듈로부터 액션 데이터가 인가되면, 기설정된 환경 데이터에 따라 인가된 액션 데이터에 대응하여 상태 데이터와 보상 데이터를 업데이트 하는 환경 모듈; 및

상기 에이전트 모듈 및 상기 환경 모듈에 미리 설정되거나 업데이트 되는 데이터 중 적어도 하나를 인가받아 기지정된 화면으로 구성하여 출력하고, 사용자에 의해 상기 에이전트 모듈 및 상기 환경 모듈에 미리 설정되거나 업데이트되는 데이터 중 적어도 하나의 데이터를 변경하기 위한 변경 데이터가 설정되어 인가되면, 인가된 변경 데이터에 대응하는 데이터를 변경 데이터로 대체하는 환경 변경 모듈을 포함하되,

상기 환경 변경 모듈은

상기 변경 데이터의 종류에 따라 기지정된 차원을 확인하고, 상기 변경 데이터의 차원이 기지정된 차원과 상이하면, 상기 변경 데이터의 차원이 기지정된 차원이 되도록 변환하여, 상기 에이전트 모듈 또는 상기 환경 모듈로 인가하는 인공지능 학습 장치.

청구항 2

삭제

청구항 3

제1항에 있어서, 상기 환경 변경 모듈은

상기 변경 데이터의 차원이 기지정된 차원이 되도록 제로 패딩 기법에 따라 변환하는 인공지능 학습 장치.

청구항 4

제1항에 있어서, 상기 환경 변경 모듈은

상기 에이전트 모듈 및 상기 환경 모듈로부터 미리 설정되거나 업데이트 되는 데이터가 저장된 메모리 어드레스를 인가받아, 메모리에서 미리 설정되거나 업데이트 되는 데이터를 획득하는 인공지능 학습 장치.

청구항 5

제4항에 있어서, 상기 환경 변경 모듈은

상기 변경 데이터가 인가되면, 상기 메모리에 변경 데이터를 저장하고, 저장된 변경 데이터의 메모리 어드레스를 상기 에이전트 모듈 또는 상기 환경 모듈로 인가하는 인공지능 학습 장치.

청구항 6

인공 지능 에이전트 모듈을 학습시키기 위한 컴퓨팅 장치에서 수행되는 방법으로서,

상기 에이전트 모듈과 상기 에이전트 모듈이 액션 데이터를 출력할 수 있도록 상태 데이터와 보상 데이터를 제공하는 환경 모듈을 구동시키기 위한 데이터를 설정하는 단계;

설정된 데이터에 따라 상기 에이전트 모듈과 상기 환경 모듈을 구동하여 강화 학습을 수행하는 단계; 및

강화 학습 중 사용자에 의해 상기 에이전트 모듈 및 상기 환경 모듈에 미리 설정되거나 업데이트되는 데이터 중 적어도 하나의 데이터를 변경하기 위한 변경 데이터가 설정되어 인가되면, 인가된 변경 데이터에 대응하는 데이터를 변경 데이터로 대체하는 단계를 포함하되,

상기 변경 데이터로 대체하는 단계는

변경 데이터가 설정되어 인가되면, 인가된 상기 변경 데이터의 종류에 따라 기지정된 차원을 확인하는 단계;

상기 변경 데이터의 차원이 기지정된 차원과 상이하면, 상기 변경 데이터의 차원이 기지정된 차원이 되도록 변환하는 단계; 및

변환된 변경 데이터를 상기 에이전트 모듈 또는 상기 환경 모듈로 인가하는 단계를 포함하는 인공지능 학습 방법.

청구항 7

삭제

청구항 8

제6항에 있어서, 상기 변환하는 단계는

상기 변경 데이터의 차원이 기지정된 차원이 되도록 제로 패딩 기법에 따라 변환하는 인공지능 학습 방법.

청구항 9

제6항에 있어서, 상기 데이터를 설정하는 단계는

상기 에이전트 모듈 및 상기 환경 모듈로부터 미리 설정되거나 업데이트 되는 데이터가 저장된 메모리 어드레스를 인가받는 단계; 및

메모리에서 미리 설정되거나 업데이트 되는 데이터를 획득하는 인공지능 학습 방법.

청구항 10

제9항에 있어서, 상기 변경 데이터로 대체하는 단계는

상기 변경 데이터가 인가되면, 상기 메모리에 변경 데이터를 저장하는 단계; 및

저장된 변경 데이터의 메모리 어드레스를 상기 에이전트 모듈 또는 상기 환경 모듈로 인가하는 단계를 포함하는 인공지능 학습 방법.

발명의 설명

기 술 분 야

[0001] 본 발명은 인공지능 학습 장치 및 방법에 관한 것으로, 강화 학습에 기반하는 학습 수행 중 액터, 액션, 보상 및 상태 등을 신규하게 변경할 수 있는 인공지능 학습 장치 및 방법에 관한 것이다.

배 경 기 술

[0002] 강화학습은 현재의 상태(State)에서 어떤 액션(Action)을 취하는 것이 최적인지를 학습하는 기법으로, 강화 학습에서는 액션을 취할 때마다 환경(Environment)의 상태 변화에 따른 보상(Reward)이 주어지며 이러한 보상을 최대화하는 방향으로 학습이 진행된다.

[0003] 도 1은 강화 학습에 기반한 인공지능 학습 방법의 개념을 설명하기 위한 도면이다.

[0004] 도 1을 참조하면, 강화 학습에서는 학습 대상이 되는 에이전트 모듈(110)과 에이전트 모듈(110)을 학습시키기 위한 환경 모듈(120)로 구성될 수 있다.

[0005] 에이전트 모듈(110)은 환경 모듈(120)로부터 현재 상태(S(t))와 이전 액션(A(t-1))에 대한 보상(R(t))이 주어지면, 현재 상태(S(t))에서 보상(R(t))이 더 증가되도록 액션(A(t))을 결정한다. 그리고 환경 모듈(120)은 에이전트 모듈(110)에서 결정한 액션(A(t))에 따라 상태(S(t))를 다음 상태(S(t+1))로 업데이트하고, 업데이트된 상태(S(t+1))에 따른 보상(R(t+1))을 판단한다.

[0006] 이와 같이 강화 학습에서는 에이전트 모듈(110)이 선택하는 액션에 따라 변화되는 상태와 보상을 지속적으로 반복하여 반영함으로써, 이후 더 나은 액션을 선택하도록 학습되는 학습 기법을 일컫는다.

[0007] 그러나 기존의 강화 학습에서는 초기 액터와 초기 환경, 초기 상태, 초기 보상 등과 같은 초기 데이터가 설정된 이후로는 반복되는 강화 학습에 의해 초기 설정된 데이터가 반복적으로 업데이트될 뿐, 새로운 데이터를 입력할 수 있는 수단이 제공되지 않았다. 이에 기존에는 강화 학습 도중에 데이터가 완전히 새로운 데이터로 변경될 수 없었다. 만일 새로운 데이터로 강화 학습을 수행하기 위해서는 이전 수행하던 강화 학습을 종료시켜 학습된 내용을 모두 초기화한 이후 다시 새로운 데이터를 초기 데이터로 인가하여 새로이 강화 학습을 수행해야 하였다.

[0008] 특히 새로이 설정하고자 하는 데이터의 차원(dimension 또는 차수)가 이전 설정되거나 업데이트된 데이터와 상이한 경우에는 차원의 차이로 인해 설정하고자 하는 데이터를 입력할 수 있는 방법이 없었다는 한계가 있다.

[0009] 그러나 최근 강화 학습 기법이 적용되는 분야가 다양해짐에 따라 강화 학습 중인 에이전트 모듈과 환경 모듈에 다른 차원의 데이터로 변경하고자 하는 요구가 증대되고 있다. 일 예로 에이전트 모듈(110)의 액터가 8개의 다리를 갖는 거미 로봇이고, 강화 학습 기법에 따라 거미 로봇의 동작을 학습시키는 경우를 가정하면, 기존에는 에이전트 모듈(110)이 액터인 거미 로봇에 설정된 8개의 다리를 주변 환경에 따라 최적으로 구동하는 액션만을 학습하게 된다. 그러나 다양한 환경 요인에 의해 8개의 다리 중 하나 또는 그 이상의 다리가 구동 불능인 상태가 발생할 수도 있으며, 이 경우 기존의 강화 학습 방식으로 학습된 거미 로봇은 7개 또는 그 이하의 다리를 구동할 수 없게 되는 문제가 발생하게 된다. 즉 강화 학습된 에이전트 모듈이 차원의 변화에 대처하지 못하는 한계가 있다.

선행기술문헌

특허문헌

[0010] (특허문헌 0001) 한국 등록 특허 제10-2079745호 (2020.02.14 등록)

발명의 내용

해결하려는 과제

[0011] 본 발명의 목적은 강화 학습 중에 각종 데이터를 다양하게 변화시킬 수 있는 인공지능 학습 장치 및 방법을 제공하는데 있다.

[0012] 본 발명의 다른 목적은 초기 설정값과 다른 설정값으로도 강화 학습을 계속 수행할 수 있도록 하는 인공지능 학습 장치 및 방법을 제공하는데 있다.

과제의 해결 수단

[0013] 상기 목적을 달성하기 위한 본 발명의 일 실시예에 따른 인공지능 학습 장치는 상태 데이터와 보상 데이터가 인가되면, 기설정된 액터 데이터에 따라 액션 데이터를 출력하는 에이전트 모듈; 상기 에이전트 모듈로부터 액션 데이터가 인가되면, 기설정된 환경 데이터에 따라 인가된 액션 데이터에 대응하여 상태 데이터와 보상 데이터를 업데이트 하는 환경 모듈; 및 상기 에이전트 모듈 및 상기 환경 모듈에 미리 설정되거나 업데이트 되는 데이터 중 적어도 하나를 인가받아 기지정된 화면으로 구성하여 출력하고, 사용자에 의해 상기 에이전트 모듈 및 상기 환경 모듈에 미리 설정되거나 업데이트되는 데이터 중 적어도 하나의 데이터를 변경하기 위한 변경 데이터가 설정되어 인가되면, 인가된 변경 데이터에 대응하는 데이터를 변경 데이터로 대체하는 환경 변경 모듈을 포함한다.

[0014] 상기 환경 변경 모듈은 상기 변경 데이터의 종류에 따라 기지정된 차원을 확인하고, 상기 변경 데이터의 차원이 기지정된 차원과 상이하면, 상기 변경 데이터의 차원이 기지정된 차원이 되도록 변환하여, 상기 에이전트 모듈 또는 상기 환경 모듈로 인가할 수 있다.

[0015] 상기 환경 변경 모듈은 상기 변경 데이터의 차원이 기지정된 차원이 되도록 제로 패딩 기법에 따라 변환할 수 있다.

[0016] 상기 환경 변경 모듈은 상기 에이전트 모듈 및 상기 환경 모듈로부터 미리 설정되거나 업데이트 되는 데이터가 저장된 메모리 어드레스를 인가받아, 메모리에서 미리 설정되거나 업데이트 되는 데이터를 획득할 수 있다.

[0017] 상기 환경 변경 모듈은 상기 변경 데이터가 인가되면, 상기 메모리에 변경 데이터를 저장하고, 저장된 변경 데이터의 메모리 어드레스를 상기 에이전트 모듈 또는 상기 환경 모듈로 인가할 수 있다.

[0018] 상기 목적을 달성하기 위한 본 발명의 다른 실시예에 따른 인공지능 학습 방법은 상기 에이전트 모듈과 상기 에이전트 모듈이 액션 데이터를 출력할 수 있도록 상태 데이터와 보상 데이터를 제공하는 환경 모듈을 구동시키기 위한 데이터를 설정하는 단계; 설정된 데이터에 따라 상기 에이전트 모듈과 상기 환경 모듈을 구동하여 강화 학습을 수행하는 단계; 및 강화 학습 중 사용자에 의해 상기 에이전트 모듈 및 상기 환경 모듈에 미리 설정되거나 업데이트되는 데이터 중 적어도 하나의 데이터를 변경하기 위한 변경 데이터가 설정되어 인가되면, 인가된 변경 데이터에 대응하는 데이터를 변경 데이터로 대체하는 단계를 포함한다.

발명의 효과

[0019] 따라서, 본 발명의 실시예에 따른 인공지능 학습 장치 및 방법은 강화 학습 중에 액터, 액션, 보상 및 상태 등을 다양하게 변화시킬 수 있을 뿐만 아니라, 다른 차원의 값으로 변화시킬 수 있도록 하여, 각종 예기치 못한 환경 변화에도 유연하게 대응할 수 있도록 학습시킬 수 있다.

도면의 간단한 설명

[0020] 도 1은 강화 학습에 기반한 인공지능 학습 방법의 개념을 설명하기 위한 도면이다.

도 2는 본 발명의 일 실시예에 따른 인공지능 학습 장치로 구현될 수 있는 컴퓨팅 장치의 예를 나타낸다.

도 3은 본 발명의 일 실시예에 따른 인공지능 학습 장치의 개략적 구조를 나타낸다.

도 4는 도 3의 인공지능 학습 장치에서 환경 변경 모듈의 상세 구성의 일 예를 나타낸다.

도 5는 환경 변경 인터페이스의 일 예를 나타낸다.

도 6 및 도 7은 도 3의 데이터 설정부가 차원이 변경되어 인가된 데이터를 처리하는 방식을 설명하기 위한 도면이다.

도 8은 본 발명의 일 실시예에 따른 인공지능 학습 방법을 나타낸다.

발명을 실시하기 위한 구체적인 내용

[0021] 본 발명과 본 발명의 동작상의 이점 및 본 발명의 실시에 의하여 달성되는 목적을 충분히 이해하기 위해서는 본 발명의 바람직한 실시예를 예시하는 첨부 도면 및 첨부 도면에 기재된 내용을 참조하여야만 한다.

[0022] 이하, 첨부한 도면을 참조하여 본 발명의 바람직한 실시예를 설명함으로써, 본 발명을 상세히 설명한다. 그러나, 본 발명은 여러 가지 상이한 형태로 구현될 수 있으며, 설명하는 실시예에 한정되는 것이 아니다. 그리고, 본 발명을 명확하게 설명하기 위하여 설명과 관계없는 부분은 생략되며, 도면의 동일한 참조부호는 동일한 부재임을 나타낸다.

[0023] 명세서 전체에서, 어떤 부분이 어떤 구성요소를 "포함"한다고 할 때, 이는 특별히 반대되는 기재가 없는 한 다른 구성요소를 제외하는 것이 아니라, 다른 구성요소를 더 포함할 수 있는 것을 의미한다. 또한, 명세서에 기재된 "...부", "...기", "모듈", "블록" 등의 용어는 적어도 하나의 기능이나 동작을 처리하는 단위를 의미하며, 이는 하드웨어나 소프트웨어 또는 하드웨어 및 소프트웨어의 결합으로 구현될 수 있다.

[0024] 도 2는 본 발명의 일 실시예에 따른 인공지능 학습 장치로 구현될 수 있는 컴퓨팅 장치의 예를 나타낸다.

[0025] 도 2를 참조하면, 본 발명의 일 실시 예에 따른 컴퓨팅 장치는 프로세서(210) 메모리(220) 및 통신부(230)를 포함할 수 있다.

[0026] 프로세서(210)는 MPU(micro processing unit), CPU(central processing unit)등으로 구현되어 인공지능 학습 장치로서 동작할 수 있다. 메모리(220)는 프로세서(210)에서 동작을 수행하기 위한 각종 데이터를 저장하여, 프로세서(210)로 저장된 데이터를 전달하거나, 프로세서(210)에서 인가되는 데이터를 저장한다. 여기서는 예시로서 메모리(220)를 프로세서(210)와 별도의 구성 요소로 표시하였으나, 메모리(220)는 프로세서(210)에 포함되어 구성될 수도 있다. 일 예로 메모리(220)는 프로세서(210) 내에 포함되는 캐시 메모리로 구현될 수도 있다.

통신부(230)는 외부의 장치 또는 프로세서(210) 외부에 구현되는 모듈 등과 통신을 수행하여 외부로 데이터를 전송하거나 외부의 데이터를 인가받아 프로세서(210)로 전달할 수 있다.

[0027] 도 3은 본 발명의 일 실시예에 따른 인공지능 학습 장치의 개략적 구조를 나타낸다.

[0028] 도 3을 참조하면, 본 실시예에 따른 인공지능 학습 장치는 도 1과 마찬가지로 에이전트 모듈(310)과 환경 모듈(320)을 포함한다. 다만 본 실시예에 따른 인공지능 학습 장치는 환경 변경 모듈(330)을 더 포함한다. 여기서 에이전트 모듈(310)과 환경 모듈(320) 및 환경 변경 모듈(330)은 도 2의 프로세서(210) 내의 하드웨어로 구현되거나, 프로세서(210)에서 실행되는 소프트웨어 모듈로 구현될 수 있다.

[0029] 에이전트 모듈(310)은 강화 학습 대상으로서, 환경 모듈(320)로부터 현재 상태(S(t))와 이전 액션(A(t-1))에 대한 보상(R(t))이 주어지면, 현재 상태(S(t))에서 보상(R(t))이 더 증가되도록 액션(A(t))을 결정하여 환경 모듈(320)로 출력한다. 에이전트 모듈(310)은 메모리(220)에 저장된 액터 데이터를 인가받고, 인가된 액터 데이터를 기반으로 현재 상태(S(t))와 보상(R(t))에 대응하는 액션(A(t))을 결정할 수 있다. 여기서 결정된 액션(A(t))은 메모리(220)에 저장된다. 또한 에이전트 모듈(310)은 현재 상태(S(t))와 보상(R(t))에 대한 데이터 또한 환경 모듈(320)로부터 직접 인가받는 것이 아니라, 환경 모듈(320)이 메모리(220)에 저장한 상태(S(t))와 보상(R(t))을 읽어와서 액션(A(t))을 결정할 수 있다.

[0030] 그리고 환경 모듈(320)은 에이전트 모듈(310)이 메모리(220)에 저장한 액션(A(t))을 읽어 확인하고, 확인된 액션(A(t))에 따라 상태(S(t))를 다음 상태(S(t+1))로 업데이트하여 메모리(220)에 저장한다. 또한 업데이트된 상태(S(t+1))에 따른 보상(R(t+1))을 판단하여 메모리(220)에 저장한다. 이때, 환경 모듈(120)은 액션(A(t))에 대한 보상을 즉시 반영하지 않고, 지연하여 반영할 수도 있다.

[0031] 환경 변경 모듈(330)은 항시 구동될 수도 있으나, 사용자의 요청에 응답하여 구동되는 것이 바람직하다. 환경 변경 모듈(330)은 에이전트 모듈(310)과 환경 모듈(320)로부터 액터(actor), 환경, 상태(S(t)), 보상(R(t)) 등의 다양한 데이터를 획득할 수 있다. 이때, 환경 변경 모듈(330)은 에이전트 모듈(310)과 환경 모듈(320)로부터 직접 데이터를 획득할 수도 있으나, 메모리(220)에 저장한 데이터를 획득할 수도 있다.

[0032] 그리고 환경 변경 모듈(330)은 획득된 데이터를 기지정된 방식으로 출력하여 사용자에게 표시한다. 환경 변경 모듈(330)은 도 2의 통신부(230)를 통해 디스플레이 장치(미도시)로 획득한 데이터가 출력되도록 하거나 다른 컴퓨팅 장치로 획득한 데이터를 전송하여 출력할 수도 있다.

[0033] 그리고 사용자로부터 변경 데이터가 인가되면, 환경 변경 모듈(330)은 인가된 변경 데이터를 에이전트 모듈(310)과 환경 모듈(320)로 인가하여, 액터, 환경, 상태(S(t)), 보상(R(t))등을 변경시킨다. 이때에도 환경 변경 모듈(330)은 에이전트 모듈(310)과 환경 모듈(320)로 변경 데이터를 직접 전달하여 변경시킬 수도 있으나, 인가된 변경 데이터를 메모리(220)에 저장하고, 변경 데이터가 저장된 메모리 주소를 에이전트 모듈(310)과 환경 모듈(320)로 전달함으로써, 에이전트 모듈(310)과 환경 모듈(320)이 인가된 메모리 주소에 따라 변경 데이터를 획득하도록 할 수도 있다.

[0034] 도 4는 도 3의 인공지능 학습 장치에서 환경 변경 모듈의 상세 구성의 일 예를 나타내고, 도 5는 환경 변경 인터페이스의 일 예를 나타내며, 도 6 및 도 7은 도 3의 데이터 설정부가 차원이 변경되어 인가된 데이터를 처리하는 방식을 설명하기 위한 도면이다.

[0035] 도 4를 참조하면, 환경 변경 모듈(330)은 데이터 획득부(410), 인터페이스 제공부(420), 데이터 설정부(430) 및 데이터 변경부(440)를 포함할 수 있다.

[0036] 데이터 획득부(410)는 에이전트 모듈(310)과 환경 모듈(320)로부터 각종 데이터를 인가받아 인터페이스 제공부(420)로 전달한다. 데이터 획득부(410)는 환경 변경 모듈(330)이 구동되거나, 통신부(230)를 통해 사용자 명령으로 데이터 요청이 인가되면, 액터, 환경, 상태(S(t)), 보상(R(t)) 등의 다양한 데이터를 획득하여 인터페이스 제공부(420)로 전달할 수 있다.

[0037] 이때 데이터 획득부(410)는 에이전트 모듈(310)과 환경 모듈(320)로부터 직접 데이터를 인가받을 수도 있으나, 에이전트 모듈(310)과 환경 모듈(320)에 의해 메모리(220)에 저장된 데이터를 인가받을 수도 있다.

[0038] 그리고 데이터 획득부(410)는 통신부(230)를 통해 사용자가 설정한 변경 데이터를 인가받아 데이터 설정부(430)로 전달한다. 또는 데이터 획득부(410)는 인가된 변경 데이터를 메모리(220)에 전달하여 저장할 수 있다.

[0039] 인터페이스 제공부(420)는 데이터 획득부(410)에서 획득한 각종 데이터를 인가받아 기지정된 형식으로 데이터

변경 화면을 구성하여 통신부(230)로 전달할 수 있다. 인터페이스 제공부(420)는 사용자가 데이터를 용이하게 인식할 수 있는 다양한 형태로 데이터 변경 화면을 구성할 수 있다. 여기서 데이터 변경 화면은 현재 데이터를 사용자에게 표시할 뿐만 아니라 사용자가 변경 데이터를 입력할 수 있도록 일 예로 도 5와 같은 화면을 구성될 수 있다.

[0040] 또한 외부의 장치가 직접 변경 화면을 구성할 수 있는 경우, 인터페이스 제공부(420)는 생략될 수도 있다.

[0041] 데이터 설정부(430)는 데이터 획득부(410)로부터 변경 데이터를 인가받아, 에이전트 모듈(310)과 환경 모듈(320)이 처리할 수 있는 형식으로 변환 설정한다. 여기서는 일 예로 액터, 환경, 상태(S(t)), 보상(R(t))에 대한 데이터를 변경할 수 있는 데이터인 것으로 가정하였으므로, 데이터 설정부(430)는 일 예로 액터 설정부(431), 상태 설정부(432), 환경 설정부(433), 및 보상 설정부(434)를 포함할 수 있다. 그러나 변경할 수 있는 변경 데이터의 종류에 따라 데이터 설정부(430)의 구성은 다양하게 조절될 수 있다. 그리고 액터 설정부(431), 상태 설정부(432), 환경 설정부(433), 및 보상 설정부(434) 각각은 인가된 변경 데이터에서 대응하는 변경 데이터를 인가받아 기지정된 형식으로 변환한다. 특히 각 종류별 변경 데이터의 차원을 기지정된 차원이 되도록 변환할 수 있다. 예로서 액터 설정부(431)는 기존에 액터 데이터가 거미 로봇의 8개의 다리 각각에 대한 8차원으로 구성되는 반면, 액터 변경 데이터가 7개의 다리를 갖는 거미 로봇에 대응하여 7차원으로 구성되는 경우, 7차원의 액터 변경 데이터를 8차원으로 변환하여 출력할 수 있다.

[0042] 도 6의 (a)와 (b)는 각각 상태 데이터(S)와 액션 데이터(A)의 차원을 변환하는 방법의 일 예를 나타낸다. 도 6에 도시된 바와 같이, 데이터 설정부(430)에는 각 종류별 데이터에서 설정될 수 있는 최대 차원 크기(여기서는 일 예로 5)가 미리 지정되며, 변경 데이터가 인가되면, 인가된 변경 데이터의 종류에 따라 지정된 최대 차원 크기로 변경 데이터의 차원을 변환한다. 이때, 최대 차원 크기보다 작은 크기의 변경 데이터는 부족한 차원을 0으로 채울 수 있다. 즉 제로 패딩 기법을 적용하여 변경 데이터의 차원을 통일시킬 수 있다.

[0043] 다만 각 데이터의 종류에 따라 최대 차원 크기는 서로 상이할 수 있다. 비록 도 6에서는 상태 데이터(S)와 액션 데이터(A)가 모두 5차원인 경우를 가정하였으나, 상태 데이터(S)는 3차원인 반면, 액션 데이터(A)는 7차원으로 설정될 수도 있다. 이에 각 데이터 종류에 따라 서로 다르게 설정되는 차원으로 변경 데이터를 변환할 수 있도록 도 4에서는 데이터 설정부(430)가 액터 설정부(431), 상태 설정부(432), 환경 설정부(433), 및 보상 설정부(434)를 포함하는 것으로 도시하였다.

[0044] 한편, 데이터 설정부(430)는 인가된 변경 데이터의 차원을 직접 변경하지 않고 메모리 주소를 이용하는 간접 접근 방식을 이용할 수도 있다. 도 6에 도시된 바와 같이, 변경 데이터의 차원을 변환하는 것은 에이전트 모듈(310)과 환경 모듈(320)이 변화하는 가변 차원의 입력을 인가받지 못하도록 구성되기 때문이다. 그러나 에이전트 모듈(310)과 환경 모듈(320)이 데이터를 직접 인가받지 않고, 메모리(220)에 저장된 데이터를 인가받도록 구성된다면, 에이전트 모듈(310)과 환경 모듈(320)은 항시 동일한 형식의 메모리 주소를 획득하여 다양한 차원의 변경 데이터를 획득할 수 있다.

[0045] 도 7에서도 도 6에서와 마찬가지로 상태 데이터(S)와 액션 데이터(A)가 다양한 차원의 변경 데이터로 인가되는 경우를 도시하였다. 그러나 도 7에서는 에이전트 모듈(310)과 환경 모듈(320)이 변경 데이터를 직접 인가받지 않고, 변경 데이터의 메모리 주소를 인가받는 방식으로 구현됨에 따라 에이전트 모듈(310)과 환경 모듈(320)에는 변경 데이터의 차원에 무관하게 항시 동일 형식의 메모리 주소가 인가될 수 있다. 그리고 에이전트 모듈(310)과 환경 모듈(320)은 인가된 메모리 주소를 기반으로 다양한 차원의 변경 데이터를 읽어올 수 있다. 여기서는 이와 같이 메모리 주소를 활용하는 방식을 간접 접근 방식이라 한다.

[0046] 간접 접근 방식에서 일 예로 액터 설정부(431)는 데이터 획득부(410)가 메모리(220)에 저장한 메모리 주소를 확인하고, 확인된 메모리 주소를 데이터 변경부(440)로 전달할 수 있다.

[0047] 한편, 데이터 변경부(440)는 데이터 설정부(430)에서 변환된 변경 데이터를 에이전트 모듈(310)과 환경 모듈(320)로 직접 전달할 수 있다. 또는 데이터 설정부(430)에서 확인된 메모리 주소를 에이전트 모듈(310)과 환경 모듈(320)로 전달할 수 있다.

[0048] 환경 변경 모듈(330)은 이전까지 수행한 강화 학습에 의해 획득된 데이터 전체를 사용자가 설정한 변경 데이터에 따라 대체할 수도 있다. 다만 이 경우, 이전까지 강화 학습을 수행한 학습 결과가 모두 소실된다. 그러나 사용자는 일부 종류의 데이터만을 지정하여 변경 데이터로 변경할 수도 있다. 이 경우에는 변경 데이터로 변경된 데이터 이외에는 이전까지 수행된 강화 학습에 의해 업데이트된 데이터가 그대로 유지될 수 있다. 따라서 결과적으로 이전 강화 학습이 수행된 학습 내용을 그대로 보존하면서 추가적으로 변경된 데이터에 대한 학습이

더 수행될 수 있게 된다. 또한 데이터를 변경 데이터로 대체하는 시점이 지정될 필요가 없다. 즉 수행되던 강화 학습이 완전히 종료되지 않은 상태에서도 언제든지 데이터를 변경 데이터로 대체할 수 있어, 에이전트를 다양한 조건 환경에서 강화 학습시킬 수 있다.

[0049] 도 8은 본 발명의 일 실시예에 따른 인공지능 학습 방법을 나타낸다.

[0050] 도 2 내지 도 7을 참조하여, 도 8의 인공지능 학습 방법을 설명하면, 우선 에이전트 모듈(310)과 환경 모듈(320)을 강화 학습시키기 위한 초기 데이터를 설정한다(S10). 그리고 설정된 데이터에 따라 에이전트 모듈(310)과 환경 모듈(320)에 대해 기지정된 방식으로 강화 학습을 수행한다(S20). 이후 환경 변경 모듈이 구동되는지 판별한다(S30). 또는 환경 변경 모듈이 구동된 상태에서 데이터 요청이 인가되는지 판별할 수도 있다.

[0051] 환경 변경 모듈이 구동되거나, 데이터 요청이 인가되면, 현재까지 강화 학습에 의해 업데이트된 데이터를 획득하여 출력한다(S40). 여기서 데이터는 에이전트 모듈(310)과 환경 모듈(320)로부터 직접 획득하거나, 에이전트 모듈(310)과 환경 모듈(320)이 메모리에 저장한 데이터를 읽어서 획득할 수 있다. 또한 출력되는 데이터는 통신부(230)를 통해 사용자에게 기지정된 형식의 화면으로 표출될 수 있다.

[0052] 그리고 사용자가 설정한 변경 데이터가 입력되는지 판별한다(S50). 만일 변경 데이터가 입력되면, 입력된 변경 데이터의 차원이 기지정된 차원인지 판별한다(S60). 여기서 기지정된 차원은 각 데이터의 종류에 따라 설정될 수 있는 최대 차원일 수 있다. 만일 변경 데이터의 차원이 기지정된 차원이 아니면, 인가된 변경 데이터의 차원이 기지정된 차원이 되도록 변환한다(S70). 이때 차원 변환은 일 예로제로 패딩 기법으로 수행할 수 있다. 그리고 차원 변환된 변경 데이터를 에이전트 모듈(310)과 환경 모듈(320)로 인가하여 적용할 수 있다(S80).

[0053] 그러나 에이전트 모듈(310)과 환경 모듈(320)이 메모리(220)에 저장된 데이터를 참조하여 구동하도록 구성된 경우, 변경 데이터가 입력되면(S50), 변경 데이터를 메모리(220)에 저장하고, 변경 데이터가 저장된 메모리 주소를 확인한다. 그리고 확인된 메모리 주소를 에이전트 모듈(310)과 환경 모듈(320)로 인가하여 적용할 수도 있다(S80).

[0054] 본 발명에 따른 방법은 컴퓨터에서 실행시키기 위한 매체에 저장된 컴퓨터 프로그램으로 구현될 수 있다. 여기서 컴퓨터 판독가능 매체는 컴퓨터에 의해 액세스 될 수 있는 임의의 가용 매체일 수 있고, 또한 컴퓨터 저장 매체를 모두 포함할 수 있다. 컴퓨터 저장 매체는 컴퓨터 판독가능 명령어, 데이터 구조, 프로그램 모듈 또는 기타 데이터와 같은 정보의 저장을 위한 임의의 방법 또는 기술로 구현된 휘발성 및 비휘발성, 분리형 및 비분리형 매체를 모두 포함하며, ROM(판독 전용 메모리), RAM(랜덤 액세스 메모리), CD(컴팩트 디스크)-ROM, DVD(디지털 비디오 디스크)-ROM, 자기 테이프, 플로피 디스크, 광데이터 저장장치 등을 포함할 수 있다.

[0055] 본 발명은 도면에 도시된 실시예를 참고로 설명되었으나 이는 예시적인 것에 불과하며, 본 기술 분야의 통상의 지식을 가진 자라면 이로부터 다양한 변형 및 균등한 타 실시예가 가능하다는 점을 이해할 것이다.

[0056] 따라서, 본 발명의 진정한 기술적 보호 범위는 첨부된 청구범위의 기술적 사상에 의해 정해져야 할 것이다.

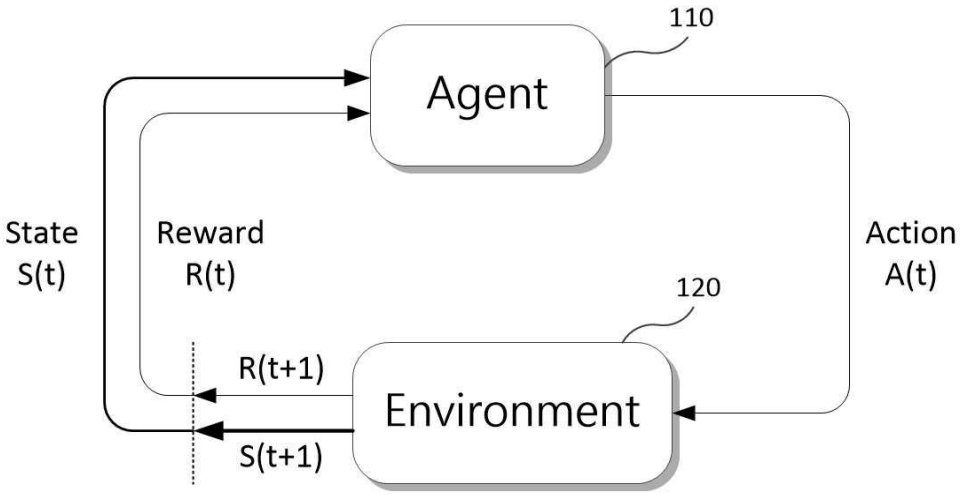
부호의 설명

[0057]

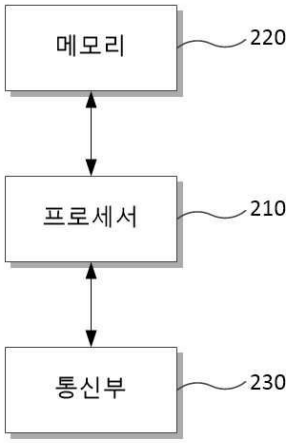
110, 310: 에이전트 모듈
120, 320: 환경 모듈
210: 프로세서
220: 메모리
230: 통신부
330: 환경 변경 모듈
410: 데이터 획득부
420: 인터페이스 제공부
430: 데이터 설정부
440: 데이터 변경부

도면

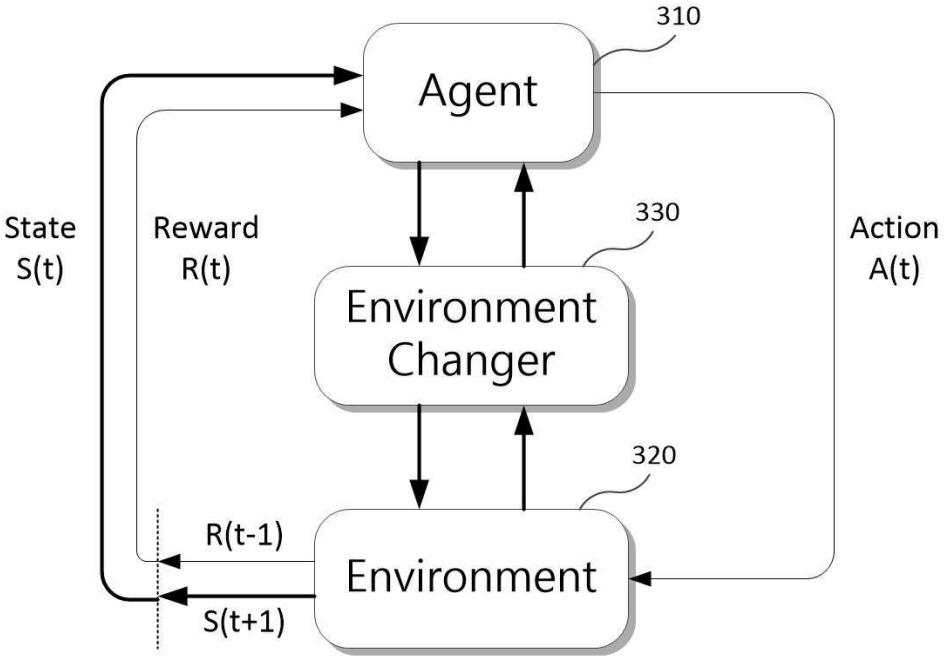
도면1


110
Agent
State
S(t)
Reward
R(t)
Action
A(t)
120
R(t+1)
Environment
S(t+1)

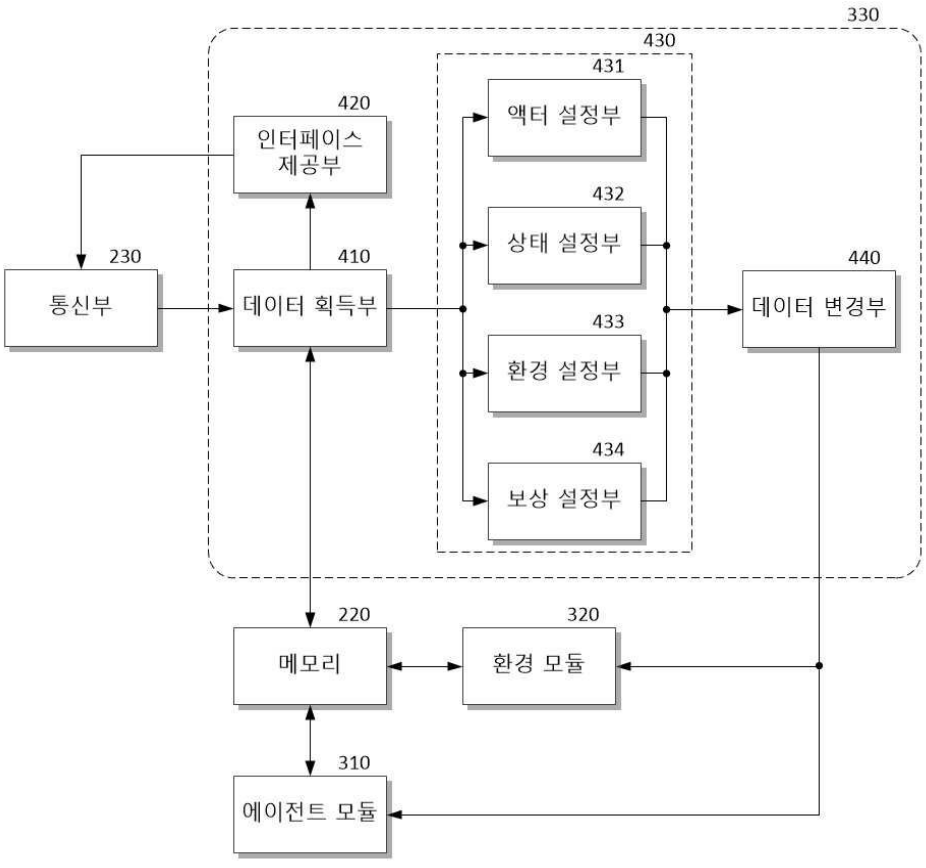
도면2


메모리
220
프로세서
210
통신부
230

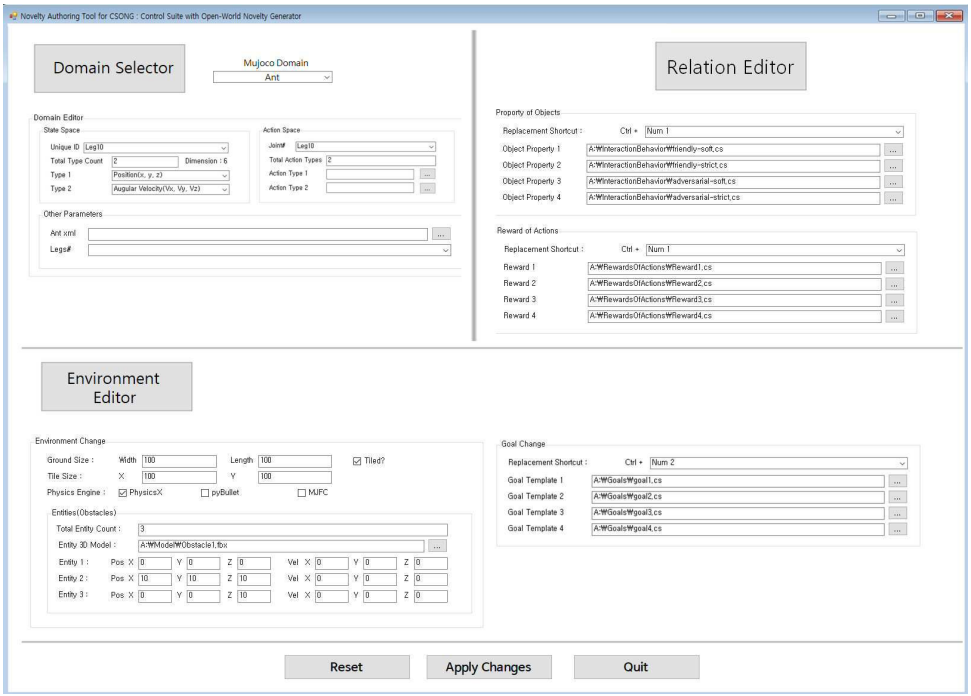
도면3


310
Agent
State
S(t)
Reward
R(t)
330
Environment
Changer
Action
A(t)
320
R(t-1)
Environment
S(t+1)

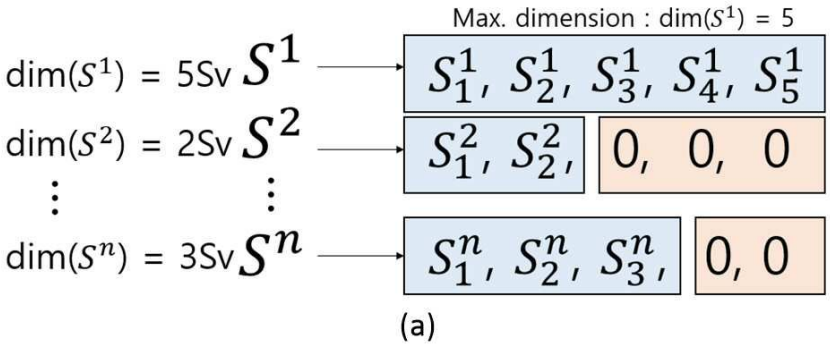
도면4


330
430
431
액터 설정부
420
인터페이스
제공부
432
상태 설정부
230
410
440
통신부
데이터 획득부
데이터 변경부
433
환경 설정부
434
보상 설정부
220
320
메모리
환경 모듈
310
에이전트 모듈

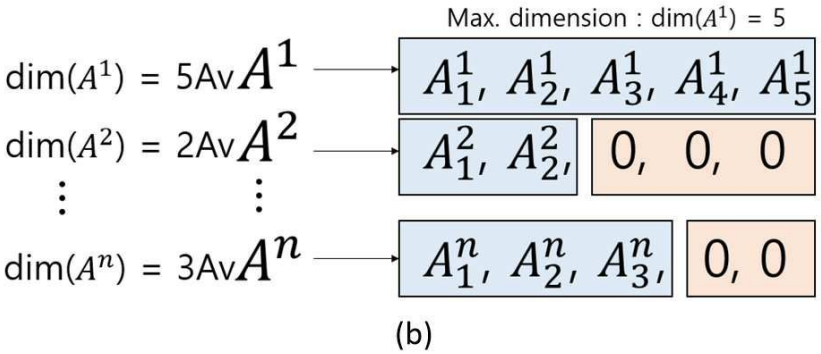
도면5


Novelty Authoring Tool for CSONG : Control Suite with Open-World Novelty Generator
Domain Selector
Mujoco Domain
Ant
Relation Editor
Domain Editor
State Space
Action Space
Other Parameters
Property of Objects
Reward of Actions
Environment Editor
Environment Change
Goal Change
Entities(Obstacles)
Reset
Apply Changes
Quit

도면6

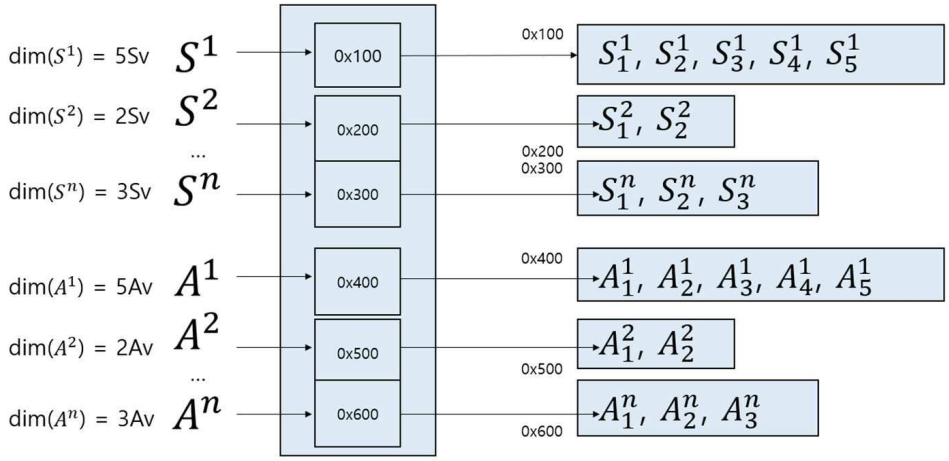

Max. dimension : dim(S^1) = 5
dim(S^1) = 5Sv S^1
dim(S^2) = 2Sv S^2
dim(S^n) = 3Sv S^n
$S_1^1, S_2^1, S_3^1, S_4^1, S_5^1$
$S_1^2, S_2^2,$ 0, 0, 0
$S_1^n, S_2^n, S_3^n,$ 0, 0

(a)

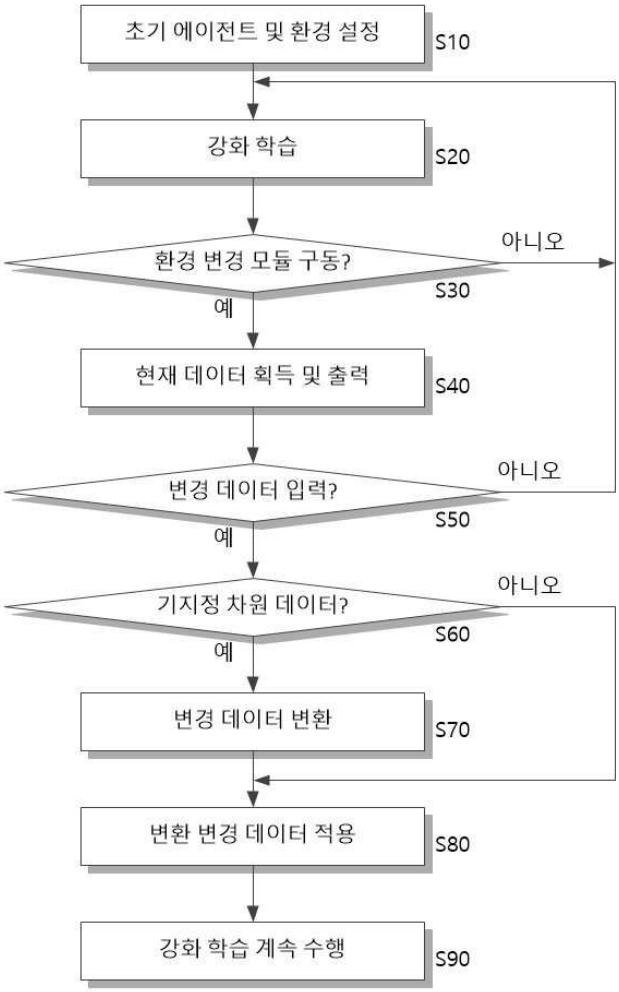

Max. dimension : dim(A^1) = 5
dim(A^1) = 5Av A^1
dim(A^2) = 2Av A^2
dim(A^n) = 3Av A^n
$A_1^1, A_2^1, A_3^1, A_4^1, A_5^1$
$A_1^2, A_2^2,$ 0, 0, 0
$A_1^n, A_2^n, A_3^n,$ 0, 0

(b)

도면7


dim(S^1) = 5Sv
S^1
dim(S^2) = 2Sv
S^2
...
dim(S^n) = 3Sv
S^n
dim(A^1) = 5Av
A^1
dim(A^2) = 2Av
A^2
...
dim(A^n) = 3Av
A^n
0x100
0x200
0x300
0x400
0x500
0x600
0x100
$S_1^1, S_2^1, S_3^1, S_4^1, S_5^1$
0x200
S_1^2, S_2^2
0x300
S_1^n, S_2^n, S_3^n
0x400
$A_1^1, A_2^1, A_3^1, A_4^1, A_5^1$
0x500
A_1^2, A_2^2
0x600
A_1^n, A_2^n, A_3^n

도면8


초기 에이전트 및 환경 설정
S10
강화 학습
S20
환경 변경 모듈 구동?
S30
아니오
예
현재 데이터 획득 및 출력
S40
변경 데이터 입력?
S50
아니오
예
기지정 차원 데이터?
S60
아니오
예
변경 데이터 변환
S70
변환 변경 데이터 적용
S80
강화 학습 계속 수행
S90