



(19) 대한민국특허청(KR)

(12) 등록특허공보(B1)

(45) 공고일자 2020년11월03일

(11) 등록번호 10-2173579

(24) 등록일자 2020년10월28일

(51) 국제특허분류(Int. Cl.)

G06N 20/00 (2019.01)

(52) CPC특허분류

G06N 20/00 (2019.01)

(21) 출원번호 10-2019-0157925

(22) 출원일자 2019년12월02일

심사청구일자 2019년12월02일

(56) 선행기술조사문헌

KR1020140063392 A

“A3C를 활용한 블록체인 기반 금융 자산 포트폴리오 관리”, 정보처리학회논문지 컴퓨터 및 통신 시스템 8권1호(pp. 17-28), 2019.01.

KR1020190069582 A

US20150100530 A1

(73) 특허권자

한국기술교육대학교 산학협력단

충청남도 천안시 동남구 병천면 충절로 1600 (한국기술교육대학교내)

(72) 발명자

한연희

대전광역시 유성구 가정로 306-6, 101동 1201호 (도룡 SK VIEW)

임현교

충청남도 천안시 동남구 병천면 가전7길 1-30, B동 304

(뒷면에 계속)

(74) 대리인

김견수

전체 청구항 수 : 총 10 항

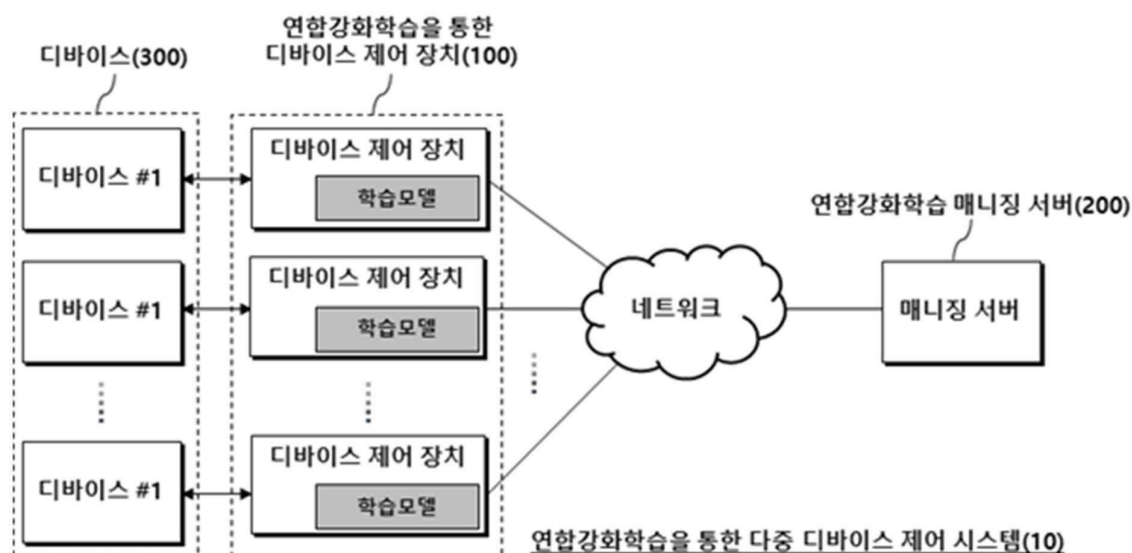
심사관 : 박승철

(54) 발명의 명칭 연합강화학습을 통한 다중 디바이스 제어 시스템 및 그 방법

(57) 요약

본 발명은 연합강화학습을 통한 다중 디바이스 제어 시스템 및 그 방법에 관한 것으로, 복수의 디바이스 제어 장치에서 복수의 디바이스를 각각 제어하기 위한 강화학습을 수행할 때, 상기 각 강화학습에 대한 그레디언트를 평균하여 상기 복수의 디바이스 제어 장치로 공유하는 그레디언트 공유(gradient sharing)과정과, 상기 그레디언트 공유과정을 통해 강화학습이 먼저 종료된 특정 디바이스 제어 장치의 학습파라미터를 상기 강화학습이 종료되지 않은 적어도 하나 이상의 디바이스 제어 장치로 전이하는 학습파라미터 전이(transfer)과정을 통해, 상기 복수의 디바이스 제어 장치에서 상기 강화학습을 연합한 연합강화학습을 수행하도록 함으로써, 상기 강화학습을 조기에 종료함과 동시에, 상기 강화학습 결과를 이용하여 상기 복수의 디바이스를 정밀하게 제어할 수 있도록 하는 연합강화학습을 통한 다중 디바이스 제어 시스템 및 그 방법에 관한 것이다.

대표도 - 도1



(72) 발명자

김찬명

충청남도 천안시 동남구 병천면 충절로 1638,
신한APT 102동 114

김주봉

충청북도 청주시 흥덕구 원터로 44, 402호

황규영

충청남도 천안시 동남구 병천면 충절로 1600, 2공
학관 433호

최호빈

충청남도 천안시 동남구 병천면 충절로 1636, 명품
에덴빌 206호

이 발명을 지원한 국가연구개발사업

과제고유번호 2018R1A6A1A03025526

부처명 교육부

과제관리(전문)기관명 National Research Foundation of Korea (한국연구재단)

연구사업명 대학중점연구소지원사업

연구과제명 상호작용 가상현실 기반 몰입형 교육 훈련 플랫폼

기 여 율 1/1

과제수행기관명 한국기술교육대학교 산학협력단

연구기간 2019.03.01 ~ 2020.02.28

공지예외적용 : 있음

명세서

청구범위

청구항 1

복수의 디바이스를 각각 제어하기 위한 강화학습을 수행하고, 상기 강화학습의 과정에서 계산되는 그래디언트와, 상기 강화학습의 완료에 따른 학습파라미터를 연합강화학습 매니징 서버로 리포팅하는 복수의 디바이스 제어 장치; 및

상기 복수의 디바이스 제어 장치로부터 리포팅된 복수의 그래디언트에 대한 평균을 계산한 평균 그래디언트를 상기 복수의 디바이스 제어 장치로 공유(sharing)하고, 상기 리포팅된 강화학습 파라미터를 상기 강화학습이 완료되지 않은 적어도 하나 이상의 디바이스 제어 장치로 전이(transfer)하는 연합강화학습 매니징 서버;를 포함하며,

상기 평균한 그래디언트를 공유하는 것과, 상기 학습파라미터를 전이하는 것을 통해 상기 복수의 디바이스 제어 장치에서, 상기 강화학습을 연합한 연합강화학습을 수행하도록 함으로써, 상기 복수의 디바이스에서 수행하는 각각의 강화학습을 완료하도록 하는 것을 특징으로 하는 연합강화학습을 통한 다중 디바이스 제어 시스템.

청구항 2

청구항 1에 있어서,

상기 복수의 디바이스 제어 장치는,

상기 연합강화학습을 통해 상기 디바이스를 제어하기 위한 학습모델을 생성하는 연합강화 학습부;를 포함하며,

상기 연합강화 학습부는,

상기 연합강화학습 매니징 서버의 요청에 따라, 현재 수행중인 강화학습에 대한 그래디언트를 계산하여 상기 연합강화학습 매니징 서버로 리포팅하는 그래디언트 리포팅부;

상기 연합강화학습 매니징 서버로부터 상기 리포팅된 복수의 그래디언트에 대한 평균을 계산한 평균 그래디언트를 수신하는 평균 그래디언트 수신부;

상기 연합강화학습 매니징 서버에 상기 학습파라미터를 리포팅하는 학습파라미터 리포팅부; 및

상기 연합강화학습 매니징 서버로부터 최초 리포팅된 상기 학습파라미터를 수신하는 학습파라미터 수신부;를 더 포함하며,

상기 연합강화 학습부는, 상기 수신한 평균 그래디언트를 이용하여 상기 강화학습을 지속하여 수행하고, 해당 강화학습이 완료되지 않은 상태에서 상기 학습파라미터가 수신되는 경우, 상기 수신한 학습파라미터를 사용하여 상기 강화학습을 수행함으로써, 상기 완료되지 않은 상태의 강화학습을 완료하기 위한 연합강화학습을 수행하는 것을 특징으로 하는 연합강화학습을 통한 다중 디바이스 제어 시스템.

청구항 3

청구항 1에 있어서,

상기 그래디언트는,

상기 강화학습을 수행하는 과정에서, 상기 강화학습이 수행되는 속도이며, 상기 평균 그래디언트를 공유하는 것을 통해서, 상기 공유한 평균 그래디언트에 따라 상기 복수의 디바이스 제어 장치를 통해 수행되는 각각의 강화학습이 진행되도록 하는 것을 특징으로 하는 연합강화학습을 통한 다중 디바이스 제어 시스템.

청구항 4

청구항 2에 있어서,

상기 복수의 디바이스 제어 장치는,

상기 생성한 학습모델을 이용하여 상기 디바이스를 제어하는 디바이스 제어부; 및

상기 디바이스를 제어한 결과인 디바이스의 상태정보를 상기 연합강화학습 매니징 서버로 제공하는 디바이스 상태정보 제공부;를 더 포함하며,

상기 복수의 디바이스 제어 장치는, 상기 연합강화학습 매니징 서버로부터, 상기 상태정보를 모니터링한 결과에 따라 상기 연합강화학습에 대한 재수행 명령이 수신되는 경우, 상기 연합강화학습을 재수행하는 것을 특징으로 하는 연합강화학습을 통한 다중 디바이스 제어 시스템.

청구항 5

청구항 1에 있어서,

상기 연합강화학습 매니징 서버는,

상기 복수의 디바이스 제어 장치에 상기 그래디언트를 요청하여 수신하는 그래디언트 수신부;

상기 복수의 디바이스 제어 장치로부터 수신한 복수의 그래디언트를 평균한 평균 그래디언트를 복수의 디바이스 제어 장치로 전송하여 공유하는 그래디언트 공유부;

상기 공유한 평균 그래디언트를 이용하여 지속한 상기 강화학습이 완료된 디바이스 제어 장치로부터, 리포팅되는 학습과라미터를 수신하는 학습과라미터 수신부; 및

상기 수신한 학습과라미터를 상기 강화학습이 완료되지 않은, 상기 적어도 하나 이상의 디바이스 제어 장치로 제공하는 학습과라미터 제공부;를 포함하는 것을 특징으로 하는 연합강화학습을 통한 다중 디바이스 제어 시스템.

청구항 6

청구항 5에 있어서,

상기 연합강화학습 매니징 서버는,

상기 복수의 디바이스 제어 장치로부터 상기 디바이스를 제어한 결과인 디바이스의 상태정보를 수신하는 디바이스 상태정보 수신부;를 더 포함하며,

상기 수신한 상태정보를 모니터링하고, 모니터링한 결과가 사전에 설정한 임계범위를 벗어나는 경우, 상기 연합강화학습에 대한 재수행 명령을 상기 복수의 디바이스 제어 장치로 전송하여 상기 연합강화학습을 재수행하도록 하는 것을 특징으로 하는 연합강화학습을 통한 다중 디바이스 제어 시스템.

청구항 7

복수의 디바이스 제어 장치에서, 디바이스를 제어하기 위한 강화학습을 개별적으로 수행하고, 연합강화학습 매니징 서버의 요청에 따라 상기 강화학습을 수행하는 과정에서의 그래디언트를 계산하여 상기 연합강화학습 매니징 서버로 리포팅하는 단계;

상기 연합강화학습 매니징 서버에서, 상기 복수의 디바이스 제어 장치로부터 리포팅된 복수의 그래디언트에 대한 평균을 계산한 평균 그래디언트를 상기 복수의 디바이스 제어 장치로 제공하여 공유하는 단계;

상기 복수의 디바이스 제어 장치에서, 상기 공유한 평균 그래디언트를 이용하여 상기 강화학습을 지속하는 단계;

상기 복수의 디바이스 제어 장치 중 적어도 하나에서, 평균 그래디언트를 이용하여 지속한 상기 강화학습이 완료되면, 상기 완료된 결과에 따른 학습과라미터를 상기 연합강화학습 매니징 서버로 리포팅하는 단계;

상기 연합강화학습 매니징 서버에서, 상기 학습과라미터를 학습이 완료되지 않은 적어도 하나 이상의 디바이스 제어 장치로 전송하여 전이하는 단계; 및

상기 적어도 하나 이상의 디바이스 제어 장치에서, 상기 연합강화학습 매니징 서버에서 전송하여 전이한 상기 학습과라미터를 사용하여 상기 강화학습을 계속하여 수행하는 단계;를 포함하며,

상기 평균 그래디언트를 공유하는 것과, 상기 학습과라미터를 전이하는 것을 통해 상기 복수의 디바이스 제어 장치에서, 상기 강화학습을 연합한 연합강화학습을 수행하도록 함으로써, 상기 복수의 디바이스 제어 장치에서

수행하는 각각의 강화학습을 완료하도록 하는 것을 특징으로 하는 연합강화학습을 통한 다중 디바이스 제어 방법.

청구항 8

청구항 7에 있어서,

상기 그래디언트는,

상기 강화학습을 수행하는 과정에서, 상기 강화학습이 수행되는 속도이며, 상기 평균 그래디언트를 공유하는 것을 통해서, 상기 공유한 평균 그래디언트에 따라 상기 복수의 디바이스 제어 장치를 통해 수행되는 각각의 강화학습이 진행되도록 하는 것을 특징으로 하는 연합강화학습을 통한 다중 디바이스 제어 방법.

청구항 9

청구항 7에 있어서,

상기 연합강화학습을 통한 다중 디바이스 제어 방법은,

상기 복수의 디바이스 제어 장치에서, 상기 연합강화학습을 통해 생성된 학습모델을 통해 상기 디바이스를 제어하는 단계; 및

상기 복수의 디바이스 제어 장치에서, 상기 디바이스를 제어한 결과인 디바이스의 상태정보를 상기 연합강화학습 매니징 서버로 제공하는 단계;를 더 포함하며,

상기 복수의 디바이스 제어 장치에서, 상기 연합강화학습 매니징 서버로부터, 상기 상태정보를 모니터링한 결과에 따라 상기 연합강화학습에 대한 재수행 명령이 수신되는 경우, 상기 연합강화학습을 재수행하는 것을 특징으로 하는 연합강화학습을 통한 다중 디바이스 제어 방법.

청구항 10

청구항 7에 있어서,

상기 연합강화학습을 통한 다중 디바이스 제어 방법은,

상기 연합강화학습 매니징 서버에서, 상기 복수의 디바이스 제어 장치로부터 상기 디바이스를 제어한 결과인 디바이스의 상태정보를 수신하는 단계;를 더 포함하며,

상기 연합강화학습 매니징 서버에서, 상기 수신한 디바이스의 상태정보를 모니터링하고, 모니터링한 결과가 사전에 설정한 임계범위를 벗어나는 경우, 상기 연합강화학습에 대한 재수행 명령을 상기 복수의 디바이스 제어 장치로 전송하여 상기 연합강화학습을 재수행하도록 하는 것을 특징으로 하는 연합강화학습을 통한 다중 디바이스 제어 방법.

발명의 설명

기술 분야

[0001]

본 발명은 연합강화학습을 통한 다중 디바이스 제어 시스템 및 그 방법에 관한 것으로, 더욱 상세하게는, 복수의 디바이스 제어 장치에서 복수의 디바이스를 각각 제어하기 위한 강화학습을 수행할 때, 상기 각 강화학습에 대한 그래디언트를 평균하여 상기 복수의 디바이스 제어 장치로 공유하는 그래디언트 공유(gradient sharing)과정과, 상기 그래디언트 공유과정을 통해 강화학습이 먼저 종료된 특정 디바이스 제어 장치의 학습파라미터를 상기 강화학습이 종료되지 않은 적어도 하나 이상의 디바이스 제어 장치로 전이하는 학습파라미터 전이(transferr)과정을 통해, 상기 복수의 디바이스 제어 장치에서 상기 강화학습을 연합한 연합강화학습을 수행하도록 함으로써, 상기 강화학습을 조기에 종료함과 동시에, 상기 강화학습 결과를 이용하여 상기 복수의 디바이스를 정밀하게 제어할 수 있도록 하는 연합강화학습을 통한 다중 디바이스 제어 시스템 및 그 방법에 관한 것이다.

배경 기술

[0002]

최근 산업기술 및 정보통신 기술의 급격한 발전으로 인해, 고유한 기능을 가지면서 다양한 업무를 처리하는 다양한 디바이스들이 개발되고 있다.

- [0003] 이러한 디바이스들은, 사람이 처리하기에 많은 시간이 소요되거나, 사람에게 위험한 업무 또는 사람이 다루기 어려운 업무를 대신하여 처리함으로써, 사람들에게 안정성과 편의성을 제공하고 있으며, 상기 디바이스들을 보다 편리하게 제어하기 위한 디바이스 제어 시스템들이 개발되고 있다.
- [0004] 그러나 종래의 디바이스 제어 시스템들은, 사용자가 직접적으로 조작하여 상기 디바이스를 제어하는 형태로 구현되기 때문에 상기 디바이스의 운용에 높은 정밀성이 요구되는 경우에는, 상기 업무의 목적을 제대로 달성하지 못하는 문제점이 발생할 수 있다.
- [0005] 이러한 문제점을 해결하기 위해 최근에는, 상기 디바이스에 인공지능을 접목하여 상기 디바이스를 보다 정확하게 정밀하게 제어할 수 있도록 하는 인공지능 기반의 디바이스 제어 시스템이 개발되어 상용화되고 있다.
- [0006] 상기 인공지능 기반의 디바이스 제어 시스템은, 기본적으로 상기 디바이스를 제어하기 위한 학습모델을 구축해야 하는데, 상기 학습모델은 상기 디바이스의 동작을 위해 생성되는 복수의 학습데이터를 학습하는 학습과정을 수행함으로써, 생성된다.
- [0007] 그러나 상기 학습모델을 구축하기 위한 학습과정은 장시간 소요되며, 적은 수의 학습데이터를 이용하여 학습을 수행하는 경우, 학습모델의 정확성이 현저하게 떨어지는 문제점이 있다.
- [0008] 따라서 복수의 디바이스 제어 장치에서, 유사한 특성과 동일한 목적을 가지고 운용되는 복수의 디바이스를 각각 제어하기 위한 강화학습을 수행할 때, 상기 강화학습을, 상기 강화학습별로 계산되는 그래디언트를 평균하여 상기 복수의 디바이스 제어 장치로 공유하는 그래디언트 공유과정과 상기 특정 디바이스 제어 장치에서 상기 그래디언트 공유과정을 통해 강화학습이 종료된 경우, 상기 강화학습이 종료된 디바이스 제어 장치의 학습파라미터를 상기 강화학습이 종료되지 않은 적어도 하나 이상의 디바이스 제어 장치로 제공하는 학습파라미터 전이과정을 포함하는 연합강화학습으로 구성하여, 상기 복수의 디바이스 제어 장치별로 수행되는 강화학습이 정확하고 신속하게 진행되도록 함으로써, 상기 복수의 디바이스를 정밀하게 제어할 수 있도록 하는 방안을 제안하고자 한다.
- [0009] 다음으로 본 발명의 기술분야에 존재하는 선행기술에 대하여 간단하게 설명하고, 이어서 본 발명이 상기 선행기술에 비해서 차별적으로 이루고자 하는 기술적 사항에 대해서 기술하고자 한다.
- [0010] 먼저 한국공개특허 제2019-0101677호(2019.09.02.)는 강화학습과 머신러닝 기법을 이용한 가속기 성능 최적화를 위한 제어 파라미터 결정 시스템에 관한 것으로, 복수개의 가속기를 제어하는 복수의 가속기 제어장치에 각각 대응하는 복수의 장치 시뮬레이터에서, 상기 가속기 제어를 위한 제어 파라미터들의 값에 대한 변경과 그에 따른 가속기 최종 출력 품질을 인공 신경망 기반 학습과정을 통해 학습하고, 상기 복수의 가속기 제어장치에서 상기 학습한 결과를 이용하여 상기 각 가속기의 최종 출력 품질이 가장 높도록 하는 제어 파라미터의 값을 산출하는 강화학습과 머신러닝 기법을 이용한 가속기 성능 최적화를 위한 제어 파라미터 결정 시스템에 관한 것이다.
- [0011] 상기 선행기술은, 인공지능 방법을 통해 복수의 가속기라는 디바이스를 제어하고자 하는 것이나, 단순히 복수의 가속기에 대한 복수의 장치 시뮬레이터를 통해, 상기 각 가속기를 제어하기 위한 학습과정을 개별적으로 수행하도록 하는 것이다.
- [0012] 반면에 본 발명은, 복수의 디바이스 제어 장치에서 강화학습을 통해 복수의 디바이스를 각각 제어하기 위한 학습모델을 생성할 때, 상기 각 강화학습 과정에서 계산되는 그래디언트를 평균하여, 상기 각 강화학습에 적용하고, 상기 강화학습이 먼저 종료된 특정 학습모델의 학습파라미터를, 상기 강화학습이 종료되지 않은 타 학습모델의 학습파라미터로 설정하도록 함으로써, 상기 모든 강화학습을 신속하게 종료함과 동시에 상기 강화학습을 통해 생성한 학습모델을 이용하여 상기 복수의 디바이스를 정밀하게 제어할 수 있도록 하는 것이다. 따라서 상기 선행기술은 이러한 본 발명의 기술적 특징을 기재하거나 시사하고 있지 않다.
- [0013] 또한 한국공개특허 제2019-0103088호(2019.09.04.)는 연합학습을 통한 단말의 명함을 인식하는 방법 및 이를 위한 장치에 관한 것으로, 명함의 이미지를 입력받는 각 단말에서, 상기 입력받은 명함의 이미지로부터 상기 명함의 주소록 필드와 관련된 텍스트가 포함된 특징값을 추출하여, 중앙의 서버에 구축된 제1 공통 예측모델에 입력하고, 상기 제1 공통 예측모델에서 출력되는 제1 텍스트정보를 상기 주소록 필드에 입력하여 사용자에게 제공함으로써, 상기 사용자로부터 상기 제1 텍스트정보에 대한 오류를 정정한 제2 텍스트정보를 제공받아, 상기 제1 텍스트정보와 상기 제2 텍스트정보를 이용하여 상기 제1 공통 예측모델을 학습시켜, 상기 학습시킨 제1 예측모델을 상기 각 단말에서 공유하여, 명함을 인식할 수 있도록 하는 연합학습을 통한 단말의 명함을 인식하는 방법 및 이를 위한 장치에 관한 것이다.

[0014] 즉, 상기 선행기술은 복수의 단말에서 제1 텍스트정보 및 제2 텍스트정보를 중앙의 서버에 각각 제공하여 상기 중앙의 서버에 구축된 제1 공통 예측모델을 학습하도록 하여 상기 제1 공통 예측모델을 공유하도록 하는 것으로, 본 발명과 같이 복수의 디바이스를 제어하기 위한 각각의 디바이스 제어 장치에서 상기 각 디바이스를 제어하기 위한 강화학습을 수행하는 과정에서 계산되는 그레디언트를 평균하여, 상기 각 디바이스 제어 장치에서 공유하도록 하여 상기 강화학습을 지속하도록 하고, 상기 강화학습이 특정 디바이스 제어 장치에서 종료된 경우, 해당 디바이스 제어 장치의 학습파라미터를 타 디바이스 제어 장치의 강화학습에 적용함으로써, 상기 디바이스 제어 장치별 강화학습을 연합한 연합강화학습을 수행하는 방법을 기재하고 있지 않아, 상기 선행기술과 본 발명은 분명한 차이점이 있다.

발명의 내용

해결하려는 과제

[0015] 본 발명은 상기와 같은 문제점을 해결하기 위해 창안된 것으로서, 유사한 특성과 동일한 목적을 가지는 복수의 디바이스에 대한 각각의 디바이스 제어 장치에서, 강화학습을 통해 상기 각 디바이스를 자동으로 제어할 수 있는 학습모델을 각각 생성하여, 상기 디바이스를 효율적이고 정밀하게 제어할 수 있도록 하는 연합강화학습을 통한 다중 디바이스 제어 시스템 및 그 방법을 제공하는 것을 그 목적으로 한다.

[0016] 또한 본 발명은, 상기 각 디바이스 제어 장치에서 상기 디바이스를 제어하기 위한 학습모델을 생성할 때 수행되는 강화학습을, 상기 현재 수행되고 있는 강화학습에서 계산되는 그레디언트를 평균하고, 상기 평균한 그레디언트를 상기 디바이스 제어 장치로 공유하는 그레디언트 공유과정과, 상기 그레디언트 공유과정을 수행한 후, 특정 디바이스 제어 장치에서의 학습과정이 종료된 경우, 해당 학습과정이 종료된 특정 디바이스 제어 장치의 학습파라미터를 타 디바이스 제어 장치로 전이하는 학습파라미터 전이과정을 포함하는 연합강화학습으로 구성함으로써, 상기 각 디바이스 제어 장치의 강화학습을 신속하게 완료함과 동시에 상기 디바이스에 대한 정밀한 제어가 가능하도록 하는 학습모델을 생성할 수 있도록 하는 연합강화학습을 통한 다중 디바이스 제어 시스템 및 그 방법을 제공하는 것을 또 다른 목적으로 한다.

[0017] 또한 본 발명은, 상기 각 디바이스 제어 장치에서 상기 생성한 학습모델을 이용하여 실제 상기 디바이스를 제어할 때, 상기 디바이스의 에이징(aging)이나, 주변 환경 상태의 변화로 인해 상기 디바이스가 제대로 제어되지 않는 경우에는, 상기 연합강화학습을 재수행하여 상기 학습모델을 업데이트함으로써, 상기 디바이스의 에이징이나 주변 환경 상태의 변화에 적응하여 상기 디바이스를 지속적으로 정밀 제어할 수 있도록 하는 연합강화학습을 통한 다중 디바이스 제어 시스템 및 그 방법을 제공하는 것을 또 다른 목적으로 한다.

과제의 해결 수단

[0018] 본 발명의 일 실시예에 따른 연합강화학습을 통한 다중 디바이스 제어 시스템은, 복수의 디바이스를 각각 제어하기 위한 강화학습을 수행하고, 상기 강화학습의 과정에서 계산되는 그레디언트와, 강화학습 종료에 따른 학습파라미터를 연합강화학습 매니징 서버로 리포팅하는 복수의 디바이스 제어 장치 및 상기 리포팅된 그레디언트를 평균하여 상기 복수의 디바이스 제어 장치로 공유하고, 상기 리포팅된 학습파라미터를 상기 강화학습이 종료되지 않은 적어도 하나 이상의 디바이스 장치로 전이하는 연합강화학습 매니징 서버를 포함하며, 상기 평균한 그레디언트를 공유하는 것과, 상기 학습파라미터를 전이하는 것을 통해 상기 복수의 디바이스 제어 장치에서, 상기 강화학습을 연합한 연합강화학습을 수행하도록 함으로써, 전체적인 강화학습을 조기에 완료하도록 하는 것을 특징으로 한다.

[0019] 또한 상기 복수의 디바이스 제어 장치는, 상기 연합강화학습을 통해 상기 디바이스를 제어하기 위한 학습모델을 생성하는 연합강화 학습부를 포함하며, 상기 연합강화 학습부는, 상기 연합강화학습 매니징 서버의 요청에 따라, 현재 수행중인 강화학습에 대한 그레디언트를 계산하여 상기 연합강화학습 매니징 서버로 리포팅하는 그레디언트 리포팅부, 상기 연합강화학습 매니징 서버로부터, 상기 리포팅된 복수의 그레디언트에 대한 평균을 계산한 평균 그레디언트를 수신하는 평균 그레디언트 수신부, 상기 연합강화학습 매니징 서버에 상기 학습파라미터를 리포팅하는 학습파라미터 리포팅부, 상기 연합강화학습 매니징 서버로부터, 최초로 리포팅된 상기 학습파라미터를 수신하는 학습파라미터 수신부를 더 포함하며, 상기 연합강화 학습부는, 상기 수신한 그레디언트의 평균을 이용하여 상기 강화학습을 지속하여 수행하고, 해당 강화학습이 종료되지 않은 상태에서 학습파라미터가 수신되는 경우, 상기 수신한 학습파라미터를 사용하여 상기 강화학습을 수행함으로써, 상기 강화학습을 조기에 종료하기 위한 연합강화학습을 수행하는 것을 특징으로 한다.

- [0020] 또한 상기 그레디언트는, 상기 강화학습을 수행하는 과정에서, 상기 강화학습이 수행되는 속도이며, 상기 그레디언트를 공유하는 것을 통해서 상기 복수의 디바이스 제어 장치를 통해 수행되는 복수의 강화학습이 평균적인 속도로 진행되도록 하는 것을 특징으로 한다.
- [0021] 또한 상기 복수의 디바이스 제어 장치는, 상기 생성한 각 학습모델을 이용하여 상기 디바이스를 제어하는 디바이스 제어부 및 상기 디바이스를 제어한 디바이스의 상태정보를 상기 연합강화학습 매니징 서버로 제공하는 디바이스 상태정보 제공부를 더 포함하며, 상기 복수의 디바이스 제어 장치는, 상기 연합강화학습 매니징 서버로부터, 상기 상태정보를 모니터링한 결과에 따라 상기 연합강화학습에 대한 재수행 명령이 수신되는 경우, 상기 연합강화학습을 재수행하는 것을 특징으로 한다.
- [0022] 또한 상기 연합강화학습 매니징 서버는, 상기 복수의 디바이스 제어 장치에 상기 그레디언트를 요청하여 수신하는 그레디언트 수신부, 상기 수신한 그레디언트를 평균한 평균 그레디언트를 복수의 디바이스 제어 장치로 전송하여 공유하는 그레디언트 공유부, 상기 공유한 그레디언트를 이용하여, 상기 강화학습이 종료된 디바이스 제어 장치로부터, 리포팅되는 학습파라미터를 수신하는 학습파라미터 수신부 및 상기 수신한 학습파라미터를 상기 강화학습이 종료되지 않은, 적어도 하나 이상의 디바이스 제어 장치로 제공하여 전이하는 학습파라미터 제공부를 포함하는 것을 특징으로 한다.
- [0023] 또한 상기 연합강화학습 매니징 서버는, 상기 복수의 디바이스 제어 장치에서, 상기 디바이스를 제어한 결과인 디바이스 상태정보를 수신하는 디바이스 상태정보 수신부를 더 포함하며, 상기 수신한 디바이스 상태정보를 모니터링하고, 모니터링한 결과가 사전에 설정한 임계범위를 벗어나는 경우, 상기 연합강화학습에 대한 재수행 명령을 상기 복수의 디바이스 제어 장치로 전송하여 상기 연합강화학습을 재수행하도록 하는 것을 특징으로 한다.
- [0024] 아울러 본 발명의 일 실시예에 따른 연합강화학습을 통한 다중 디바이스 제어 방법은, 복수의 디바이스 제어 장치에서, 디바이스를 제어하기 위한 강화학습을 개별적으로 수행하고, 연합강화학습 매니징 서버의 요청에 따라 상기 강화학습 과정에서의 그레디언트를 계산하여 상기 연합강화학습 매니징 서버로 리포팅하는 단계, 상기 연합강화학습 매니징 서버에서, 상기 리포팅된 그레디언트에 대한 평균을 계산한 평균 그레디언트를 상기 복수의 디바이스 제어 장치로 제공하여 공유하는 단계, 상기 복수의 디바이스 제어 장치에서, 상기 공유한 평균 그레디언트를 이용하여 상기 강화학습을 지속하는 단계, 상기 복수의 디바이스 제어 장치 중 적어도 하나에서, 평균 그레디언트를 이용한 강화학습이 종료되면, 상기 종료된 결과에 따른 학습파라미터를 상기 연합강화학습 매니징 서버로 리포팅하는 단계, 상기 연합강화학습 매니징 서버에서, 최초로 리포팅되어 수신되는 상기 학습파라미터를 강화학습이 종료되지 않은 상기 적어도 하나 이상의 디바이스 제어 장치로 전송하여 전이하는 단계 및 상기 적어도 하나 이상의 디바이스 제어 장치에서, 상기 수신한 학습파라미터를 사용하여 상기 강화학습을 계속하여 수행하는 단계를 포함하며, 상기 평균 그레디언트를 공유하는 것과, 상기 학습파라미터를 전송하는 것을 통해 상기 복수의 디바이스 제어 장치에서, 상기 강화학습을 연합한 연합강화학습을 수행하도록 함으로써, 전체적인 강화학습을 조기에 완료하도록 하는 것을 특징으로 한다.
- [0025] 또한 상기 그레디언트는, 상기 강화학습을 수행하는 과정에서, 상기 강화학습이 수행되는 속도이며, 상기 그레디언트를 공유하는 것을 통해서 상기 복수의 디바이스 제어 장치를 통해 수행되는 복수의 강화학습이 평균적인 속도로 진행되도록 하는 것을 특징으로 한다.
- [0026] 또한 상기 연합강화학습을 통한 다중 디바이스 제어 방법은, 상기 복수의 디바이스 제어 장치에서, 상기 연합강화학습을 통해 생성되는 학습모델을 통해 상기 디바이스를 제어하는 단계 및 상기 복수의 디바이스 제어 장치에서, 상기 디바이스를 제어한 결과에 따른 디바이스의 상태정보를 상기 연합강화학습 매니징 서버로 제공하는 단계를 더 포함하며, 상기 복수의 디바이스 제어 장치에서, 상기 연합강화학습 매니징 서버로부터, 상기 상태정보를 모니터링한 결과에 따라 상기 연합강화학습에 대한 재수행 명령이 수신되는 경우, 상기 연합강화학습을 재수행하는 것을 특징으로 한다.
- [0027] 또한 상기 연합강화학습을 통한 다중 디바이스 제어 방법은, 상기 연합강화학습 매니징 서버에서, 상기 복수의 디바이스 제어 장치로부터 상기 디바이스를 제어한 결과인 디바이스 상태정보를 수신하는 단계를 더 포함하며, 상기 연합강화학습 매니징 서버에서, 상기 수신한 디바이스 상태정보를 모니터링하고, 모니터링한 결과가 사전에 설정한 임계범위를 벗어나는 경우, 상기 연합강화학습에 대한 재수행 명령을 상기 복수의 디바이스 제어 장치로 전송하여 상기 연합강화학습을 재수행하도록 하는 것을 특징으로 한다.

발명의 효과

- [0028] 이상에서와 같이 본 발명의 연합강화학습을 통한 다중 디바이스 제어 시스템 및 그 방법은, 유사한 특성과 동일

한 목적을 가지는 복수의 디바이스를 제어하기 위한 각각의 디바이스 제어 장치에서, 상기 디바이스별로 생성한 학습데이터를 학습하는 강화학습을 통해 상기 각 디바이스의 상태정보에 따른 최적의 제어명령을 추출하기 위한 학습모델을 생성하고, 상기 생성한 학습모델을 통해 상기 디바이스 제어 장치에서 상기 디바이스를 정밀하고 효율적으로 제어할 수 있도록 하는 효과가 있다.

[0029] 또한 본 발명은, 상기 각 디바이스 제어 장치에서 수행되는 상기 강화학습을 동기화하고, 상기 각 강화학습 과정에서 산출되는 그래디언트를 평균하여 상기 각 디바이스 제어 장치로 공유하는 그래디언트 공유과정과, 상기 강화학습이 최초로 종료되는 특정 디바이스 제어 장치의 학습파라미터를 타 디바이스 제어 장치의 학습파라미터로 전이하는 학습파라미터 전이과정을 수행하여, 상기 각 디바이스 제어 장치의 강화학습을 연합한 연합강화학습을 통해 상기 강화학습을 신속하고 정확하게 종료하도록 함과 동시에 상기 강화학습을 통해 생성한 학습모델을 이용하여 상기 각 디바이스를 정밀하고 효율적으로 제어할 수 있도록 하는 효과가 있다.

도면의 간단한 설명

[0030] 도 1은 본 발명의 일 실시예에 따른 연합강화학습을 통한 다중 디바이스 제어 시스템 및 그 방법을 설명하기 위해 나타낸 개념도이다.

도 2는 본 발명의 일 실시예에 따른 연합강화학습을 통한 디바이스 제어 장치에서 수행되는 강화학습 과정을 설명하기 위해 나타낸 도면이다.

도 3은 본 발명의 일 실시예에 따른 연합강화학습을 수행하는 과정을 설명하기 위해 나타낸 도면이다.

도 4는 본 발명의 일 실시예에 따른 연합강화학습을 통한 디바이스 제어 장치의 구성을 나타낸 블록도이다.

도 5는 본 발명의 일 실시예에 따른 연합강화학습 매니징 서버의 구성을 나타낸 블록도이다.

도 6은 본 발명의 일 실시예에 따른 연합강화학습을 통한 다중 디바이스 제어하는 절차를 나타낸 흐름도이다.

발명을 실시하기 위한 구체적인 내용

[0031] 이하, 첨부한 도면을 참조하여 본 발명의 연합강화학습을 통한 다중 디바이스 제어 시스템 및 그 방법에 대한 바람직한 실시예를 상세히 설명한다. 각 도면에 제시된 동일한 참조부호는 동일한 부재를 나타낸다. 또한 본 발명의 실시예들에 대해서 특정한 구조적 내지 기능적 설명들은 단지 본 발명에 따른 실시예를 설명하기 위한 목적으로 예시된 것으로, 다르게 정의되지 않는 한, 기술적이거나 과학적인 용어를 포함해서 여기서 사용되는 모든 용어들은 본 발명이 속하는 기술분야에서 통상의 지식을 가진 자에 의해 일반적으로 이해되는 것과 동일한 의미를 가지고 있다. 일반적으로 사용되는 사전에 정의되어 있는 것과 같은 용어들은 관련 기술의 문맥상 가지는 의미와 일치하는 의미를 가지는 것으로 해석되어야 하며, 본 명세서에서 명백하게 정의하지 않는 한, 이상적이거나 과도하게 형식적인 의미로 해석되지 않는 것이 바람직하다.

[0032] 도 1은 본 발명의 일 실시예에 따른 연합강화학습을 통한 다중 디바이스 제어 시스템 및 그 방법을 설명하기 위해 나타낸 개념도이다.

[0033] 도 1에 도시한 바와 같이, 본 발명의 일 실시예에 따른 연합강화학습을 통한 다중 디바이스 제어 시스템(10)은, 복수의 디바이스(300), 상기 복수의 디바이스(300)를 자동으로 제어하기 위한 것으로, 상기 복수의 디바이스(300)와 각각 대응하는 복수의 디바이스 제어 장치(100) 및 상기 복수의 디바이스(300)와 상기 복수의 디바이스 제어 장치(100)를 관리하는 관리 서버(200)를 포함하여 구성된다.

[0034] 상기 복수의 디바이스(300)는, 특정 물건을 이송하는 무인이송수단(예: 로봇, 로봇 팔, 드론 등), 특정 물건을 제작하는 제작수단 등과 같이 유사한 특성을 가지면서 동일한 목적을 달성하기 위해 사용되는 복수의 장치를 의미한다.

[0035] 또한 상기 복수의 디바이스(300)는, 상기 각 디바이스 제어 장치(100)의 제어에 따른 제어상태에 대한 상태정보를 수집하기 위한 복수의 센서(미도시)를 구비한다. 상기 센서는, 상기 제어에 따른 디바이스의 상태정보를 수집하여 상기 디바이스 제어 장치(100)로 제공하는 기능을 수행한다.

[0036] 또한 상기 디바이스 제어 장치(100)는, 상기 디바이스(300)를 제어하기 위한 학습모델을 구비하고 있으면서, 상기 학습모델에 대한 강화학습을 수행하여, 상기 디바이스 제어 장치(100)의 기능이나 이용목적에 따른 정밀한 제어를 위한 최종적인 학습모델을 생성하게 된다.

- [0037] 상기 강화학습을 수행하기 위해 디바이스 제어 장치(100)는, 상기 디바이스(300)를 제어하기 위한 제어정보를 상기 학습모델을 통해 추출하여, 상기 디바이스(300)로 전송하며, 상기 디바이스(300)로부터 상기 제어정보에 따라 제어된 디바이스(300)의 다음 상태정보를 수신하여, 이에 대한 리워드를 생성한다.
- [0038] 이후, 상기 디바이스 제어 장치(100)는, 상기 리워드를 포함하는 강화학습 데이터를 구성하여, 상기 구성한 강화학습 데이터를 이용하여 상기 학습모델에 대한 강화학습을 수행한다. 이러한 과정을 반복적으로 수행하여 해당 학습모델을 고도화시켜, 상기 디바이스(300)의 기능이나 이용목적에 따라 정밀하게 제어할 수 있도록 하는 최종적인 학습모델을 생성하게 된다.
- [0039] 한편, 상기 구성되는 강화학습 데이터는, 상기 디바이스(300)의 현재 상태정보, 상기 현재 상태정보에 따른 제어정보, 상기 제어정보에 따라 상기 디바이스가 제어된 다음 상태정보(즉, 이전 상태정보에서 상기 제어정보에 따라 전이된 상태정보를 의미함), 상기 제어정보에 따라 상기 디바이스(300)가 다음 상태정보로 변화될 때, 사전에 설정한 임계범위 내에서 상기 디바이스(300)가 잘 동작되었는지에 대한 여부를 계산한 리워드를 포함한다. 이때, 상기 리워드는, 상기 제어정보를 토대로 제어된 상기 디바이스(300)의 다음 상태정보에 대한 보상값을 의미한다.
- [0040] 또한 상기 현재 상태정보 및 다음 상태정보는, 상기 디바이스(300)의 동작 주체가 되는 액추에이터(예: 모터)의 상태정보(예: 각도, 각속도)나, 상기 디바이스(300)가 상기 제어정보에 따라 변화되는 위치정보 등을 포함할 수 있다. 즉, 상기 현재 상태정보 및 다음 상태정보를 포함하는 디바이스 상태정보는, 상기 디바이스(300)의 고유한 기능이나 종류에 따라 다양하게 설정될 수 있다. 이와 마찬가지로 상기 제어정보 또한 상기 디바이스(300)의 기능이나 종류에 따라 다양하게 설정될 수 있다.
- [0041] 하나의 예로써, 상기 복수의 디바이스(300)가 스마트 공장에 적용되어 특정 물건을 이송하는 기능을 수행하는 무인이송수단인 경우, 상기 디바이스 상태정보는, 해당 무인이송수단의 액추에이터에 대한 각도, 각속도 또는 해당 무인이송수단의 위치정보를 포함할 수 있고, 상기 제어정보는, 상기 무인이송수단의 현재 상태정보를 토대로 상기 액추에이터를 구동하거나, 상기 무인이송수단을 이동시키도록 하는 제어명령을 포함할 수 있다. 또한 상기 리워드는, 상기 제어정보에 따라 제어된 결과인 다음 상태정보가, 사전에 설정한 임계범위에 따라 생성되는 것으로, 임계범위 내인 경우, 상기 디바이스(300)가 제대로 동작한 것으로 판단하여, 사전에 설정한 보상값(예: +1)을 상기 다음 상태정보에 적용하고, 임계범위를 벗어나는 경우, 상기 디바이스(300)가 제대로 동작하지 않은 것으로 판단하여 사전에 설정한 보상값(예: -1)을 적용한다.
- [0042] 이때, 상기 임계범위를 벗어나는 경우, 상기 리워드는 상기 임계범위에 근접하는 비율에 따라 달리 적용될 수 있다. 즉, 상기 임계범위를 크게 벗어날수록 상기 보상값의 크기(즉, 마이너스 값)를 점점 크게 하여, 상기 벗어나는 정도에 따라 달리 적용할 수 있는 것이다.
- [0043] 한편 상기 강화학습 데이터는, 상기 각 디바이스 제어 장치(100)별로 구성되는 것으로, 상기 디바이스(300)를 제어하여 실시간으로 구성되어 상기 강화학습에 이용되거나, 상기 연합강화학습을 통한 다중 디바이스 제어 시스템(10)의 관리자가 상기 디바이스 제어 장치(100)를 통해 상기 디바이스(300)를 직접 제어하여, 그 결과를 누적 저장함으로써, 사전에 미리 구성될 수 있다.
- [0044] 또한 상기 강화학습은, 상기 각 디바이스 제어 장치(100)별로 수행되는 강화학습을 연합한 연합강화학습으로 수행되는데, 상기 연합강화학습은 상기 각 디바이스 제어 장치(100)별로 수행되는 강화학습의 과정에서 계산되는 그레디언트를 공유하는 그레디언트 공유과정과, 상기 강화학습이 종료된 특정 디바이스 제어 장치(100)의 학습모델의 학습파라미터를 상기 강화학습이 종료되지 않은 다른 디바이스 제어 장치(100)로 전송하여, 상기 학습모델에 대한 강화학습에 적용하도록 하는 학습파라미터 전이(transfer)과정을 포함하여 구성된다.
- [0045] 상기 그레디언트는, 상기 학습모델의 출력인 제어정보를 통해 제어된 상기 디바이스(300)의 제어상태와, 실제 관리자가 원하는 디바이스(300)의 제어상태간의 오차를 최소화하는 강화학습과정에서, 상기 강화학습이 수행되는 속도를 의미한다.
- [0046] 또한 상기 그레디언트 공유과정은, 각 디바이스 제어 장치(100)별로 수행되는 강화학습의 과정에서 계산되는 그레디언트를 평균화하고, 상기 평균화한 그레디언트를 상기 각 디바이스 제어 장치(100)로 공유하여, 상기 각 강화학습에 이용하도록 하는 것으로, 상기 그레디언트 공유과정을 통해 복수의 디바이스 제어 장치(100)에서 수행하는 복수의 강화학습이 평균적인 속도로 진행되도록 한다.
- [0047] 또한 상기 학습파라미터 전이과정은, 상기 공유한 그레디언트(즉, 평균한 그레디언트)를 이용하여 상기 강화학

습을 지속한 결과, 특정 디바이스 제어 장치(100)에서 상기 강화학습이 종료된 경우, 상기 강화학습이 종료된 학습모델의 학습파라미터를 상기 강화학습이 종료되지 않은 다른 디바이스 제어 장치(100)로 제공하여, 상기 학습파라미터를 전이하고, 상기 강화학습이 종료되지 않은 다른 디바이스 제어 장치(100)는, 상기 제공 받은 학습파라미터를 상기 강화학습이 진행되고 있는 학습모델의 학습파라미터로 설정하여, 나머지 강화학습을 진행되도록 한다. 이를 통해, 상기 디바이스 제어 장치(100)별로 진행되는 강화학습 전체를 정확하고 조기에 완료될 수 있도록 한다.

[0048] 또한 상기 연합강화학습 매니징 서버(200)는, 상기 그래디언트 공유과정과, 상기 학습파라미터 전이과정을 통해 상기 각 디바이스 제어 장치(100)에서, 개별적으로 수행되는 강화학습을 연합한 연합강화학습을 수행하도록 제어하는 기능을 수행한다.

[0049] 이를 위해 상기 연합강화학습 매니징 서버(200)는, 상기 각 디바이스 제어 장치(100)별로 수행되는 강화학습을 동기화하고, 상기 각 강화학습의 과정에서 계산되는 상기 각 디바이스 제어 장치(100)별 그래디언트를 요청하여 수신하며, 상기 수신한 복수의 그래디언트에 대한 평균을 계산하여, 상기 계산한 그래디언트의 평균을 상기 각 디바이스 제어 장치(100)로 제공함으로써, 상기 그래디언트를 공유할 수 있도록 한다.

[0050] 또한 상기 공유한 그래디언트를 사용하여 상기 강화학습을 수행하는 디바이스 제어 장치(100)에서, 해당 강화학습이 종료되어, 상기 강화학습이 종료된 결과인 학습모델에 대한 학습파라미터가 수신되는 경우, 상기 연합강화학습 매니징 서버(200)는, 상기 수신한 학습파라미터를 상기 강화학습이 종료되지 않은 타 디바이스 제어 장치(100)로 제공하여, 전이함으로써, 상기 제공한 학습파라미터를 이용하여 나머지 강화학습을 수행할 수 있도록 한다.

[0051] 한편, 상기 디바이스 제어 장치(100)별로 수행되는 강화학습은 도 2를 참조하여 설명하도록 하며, 상기 연합강화학습은 도 3을 참조하여 상세히 설명하도록 한다.

[0052] 또한 디바이스 제어 장치(100)는, 상기 연합강화학습을 수행한 결과에 따라 생성한 학습모델을 이용하여, 상기 각 디바이스(300)의 기능이나 이용목적에 따라 개별적으로 제어하는 기능을 수행하며, 상기 제어한 결과에 따른 디바이스(300)의 현재 상태정보와 제어되기 전의 이전 상태정보를 포함하는 디바이스 상태정보를 상기 연합강화학습 매니징 서버(200)로 제공한다.

[0053] 또한 상기 연합강화학습 매니징 서버(200)는, 상기 디바이스 상태정보를 누적하여 저장하며, 사전에 설정한 임계범위 내에서 상기 디바이스(300)가 잘 동작되었는지에 대한 여부를 모니터링한다.

[0054] 이후, 상기 모니터링 결과, 사전에 설정한 임계범위 내에서 상기 디바이스 상태정보가 정밀하게 전이되지 않은 경우에는, 상기 연합학습과정을 재수행할 수 있도록 한다.

[0055] 즉, 상기 연합강화학습 매니징 서버(200)는, 상기 디바이스(300)의 사용에 따른 에이징, 또는 상기 디바이스(300)의 주변 환경이 변화되어, 상기 디바이스(300)가 정밀하게 제어되지 않은 경우에 상기 연합학습과정을 재수행하도록 함으로써, 상기 디바이스 제어 장치(100)가 상기 디바이스(300)의 에이징이나, 상기 주변 환경에 적응하여 상기 디바이스(300)를 지속적으로 정밀하게 제어할 수 있도록 하는 것이다.

[0056] 상기에서 설명한 것과 같이, 본 발명의 연합강화학습을 통한 다중 디바이스 제어 시스템(10)은, 복수의 디바이스(300)를 개별적으로 제어하기 위한 복수의 디바이스 제어 장치(100)에서 수행되는 강화학습을 연합하여, 상기 강화학습에 대한 연합강화학습을 수행할 수 있도록 함으로써, 상기 복수의 디바이스 제어 장치(100)에서 수행되는 전체 강화학습을 정확하고, 신속하게 종료할 수 있도록 함과 동시에 상기 디바이스(300)를 정밀하게 제어할 수 있도록 한다.

[0057] 도 2는 본 발명의 일 실시예에 따른 연합강화학습을 통한 디바이스 제어 장치에서 수행되는 강화학습 과정을 설명하기 위해 나타낸 도면이다.

[0058] 도 2에 도시한 바와 같이, 본 발명의 일 실시예에 따른 복수의 디바이스 제어 장치(100)는, 상기 디바이스 제어 장치(100)와 개별적으로 디바이스(300)를 각각 제어하기 학습모델에 대한 강화학습을 수행하고, 연합강화학습 매니징 서버(200)의 제어에 따라 상기 복수의 디바이스 제어 장치(100)에서 개별적으로 수행하는 강화학습을 연합하여 연합강화학습을 수행함으로써, 상기 디바이스(300)를 정밀하게 제어하기 위한 학습모델을 최종적으로 생성하게 된다.

[0059] 상기 디바이스 제어 장치(100)는, 상기 디바이스(300)의 제어를 위한 제어명령을 출력하는 학습모델에 대한 강

화학습을 수행한다.

- [0060] 상기 강화학습을 수행하기 위해 상기 디바이스 제어 장치(100)는, 상기 디바이스(300)의 현재 상태정보를 상기 학습모델에 입력하여, 상기 디바이스(300)를 제어하기 위한 제어정보를 출력하고, 상기 출력한 제어정보를 토대로 상기 디바이스(300)를 제어한다.
- [0061] 이때, 상기 디바이스 제어 장치(100)는, 상기 강화학습을 최초로 수행할 때, 상기 디바이스(300)를 초기화하여 상기 초기한 상태를 현재 상태정보로 하거나, 상기 디바이스(300)로 상태정보를 요청하여 수신하고, 상기 수신한 상태정보를 현재 상태정보로 하여 상기 학습모델에 입력함으로써, 상기 제어정보를 출력한다.
- [0062] 다음으로, 상기 디바이스 제어 장치(100)는, 상기 디바이스(300)로부터 상기 제어정보를 토대로 제어된 상태인 다음 상태정보를 수신하며, 상기 수신한 다음 상태정보에 대한 리워드를 생성한다.
- [0063] 즉, 상기 디바이스 제어 장치(100)는, 상기 수신한 디바이스(300)의 다음 상태정보에 대한 리워드를 생성하여, 상기 생성한 리워드를 포함하는 강화학습 데이터를 구성할 수 있도록 한다.
- [0064] 또한 상기 디바이스 제어 장치(100)는, 상기 디바이스(300)의 다음 상태정보에 대한 리워드가 생성된 경우, 상기 디바이스(300)의 현재 상태정보와, 상기 현재 상태정보를 토대로 추출한 제어정보, 상기 제어정보에 따라 제어된 상기 디바이스(300)의 다음 상태정보 및 상기 다음 상태정보에 대해 생성한 리워드를 포함한 강화학습 데이터를 구성한다.
- [0065] 이후, 상기 디바이스 제어 장치(100)는, 상기 구성한 강화학습 데이터를 이용하여 상기 학습모델에 대한 강화학습을 수행하며, 상기 강화학습을 반복적으로 수행하여, 상기 학습모델을 통해 상기 디바이스(300)를 정밀하게 제어할 수 있도록 한다.
- [0066] 한편, 상기 제어정보와, 제어정보에 따른 디바이스(300)의 다음 상태정보에 대한 송수신은, 낮은 전력을 소모하고 낮은 대역폭 환경에서도 고속으로 통신을 수행할 수 있도록 하는 MQTT(Message Queue Telemetry Transport) 통신방법을 통해 수행된다. 다만, 본 발명은 이에 한정하지 않으며, 5G, 이더넷 등과 같은 다양한 유무선 통신방법을 통해 상기 제어정보와 다음 상태정보를 송수신할 수 있도록 구현될 수 있다.
- [0067] 상기 학습모델에 대한 강화학습은 상기 구성한 강화학습 데이터를 이용하여 수행되며, 상기 연합강화학습을 통해 최종적으로 생성된 학습모델의 입력은, 상기 디바이스(300)의 현재 상태정보가 되며, 출력은 상기 디바이스(300)의 현재 상태에서부터 다음 상태로 전이시키기 위한 제어정보가 된다.
- [0068] 또한 상기 강화학습은, 상기 디바이스(300)의 현재 상태에서 다음 상태로 전이시키기 위한 제어정보를 추출하도록 수행된다. 더욱 상세하게 설명하자면, 상기 강화학습은 상기 디바이스(300)의 기능에 따른 최종 목적(예: 특정 물건을 특정 위치까지 이송)을 달성할 때까지 전체적인 리워드가 높은 다음 상태정보에 대한 제어명령을 추출하도록 수행된다.
- [0069] 즉, 상기 디바이스 제어 장치(100)는, 단순히 다음 상태정보에 대한 리워드가 높은 제어정보를 추출하도록 상기 학습모델에 대한 강화학습을 수행하는 것이 아니라, 상기 최종 목적에 따른 상태정보까지의 전체적인 리워드가 높은 제어정보를 추출하도록 하는 것이다.
- [0070] 도 3은 본 발명의 일 실시예에 따른 연합강화학습을 수행하는 과정을 설명하기 위해 나타낸 도면이다.
- [0071] 도 3에 도시한 바와 같이, 본 발명의 일 실시예에 따른 연합강화학습을 수행하는 과정은 우선, 복수의 디바이스 제어 장치(100)에서, 상기 연합강화학습 매니징 서버(200)의 제어에 따라, 제어 대상이 되는 디바이스(300)를 제어하기 위해 제어정보를 추출하는 학습모델에 대한 강화학습을 개별적으로 수행한다(①).
- [0072] 상기 연합강화학습 매니징 서버(200)는, 상기 각 디바이스 제어 장치(100)에서 수행되는 강화학습을 동기화하고, 상기 각 디바이스 제어 장치(100)에서, 상기 각 강화학습을 연합한 연합강화학습을 수행할 수 있도록 하여, 전체적인 강화학습을 조기 종료하고, 상기 디바이스(300)를 정밀하게 제어하는 학습모델을 생성할 수 있도록 하는 기능을 수행함은 상술한 바와 같다.
- [0073] 또한 복수의 디바이스 제어 장치(100)는, 상기 강화학습을 수행하고 있는 도 중, 상기 연합강화학습 매니징 서버(200)로부터 그레디언트를 요청명령을 수신한 경우, 상기 각 디바이스 제어 장치(100)는, 현재까지 수행한 강화학습 과정에서의 그레디언트를 계산하여 상기 연합강화학습 매니징 서버(200)로 리포팅한다.
- [0074] 다음으로, 상기 연합강화학습 매니징 서버(200)는, 상기 복수의 디바이스 제어 장치(100)로부터 각각 수신한 그

레디언트를 평균(②)하여, 상기 평균한 평균 그레디언트를 상기 각각의 연합강화학습 매니징 서버(200)로 제공함으로써, 상기 복수의 디바이스 제어 장치(100)에서 상기 그레디언트를 공유(즉, 평균 그레디언트를 공유)하는 그레디언트 공유과정을 수행하여 상기 각 디바이스 제어 장치(100)에서 상기 공유한 그레디언트를 이용하여 상기 강화학습을 지속하도록 한다.

- [0075] 다음으로, 상기 디바이스 제어 장치(100)는, 상기 평균 그레디언트를 수신한 경우, 상기 수신한 평균 그레디언트를 이용하여, 상기 강화학습을 지속한다(③).
- [0076] 여기서, 상기 그레디언트는, 상기 강화학습이 수행되는 속도를 의미하는 것으로, 상기 각 디바이스 제어 장치(100)에서 수행되는 강화학습의 속도는, 상기 디바이스 제어 장치(100)에서 개별적으로 구성되는 강화학습 데이터의 질(quality)나, 학습모델의 성능에 따라 달라질 수 있으므로, 상기 그레디언트 공유과정을 통해 상기 각 디바이스 제어 장치(100)에 의해 수행되는 복수의 강화학습이 평균적인 속도로 진행되도록 한다.
- [0077] 이후, 상기 복수의 디바이스 제어 장치(100)는, 상기 평균 그레디언트를 이용하여 상기 강화학습을 지속한 결과, 상기 학습모델이 상기 디바이스(300)의 기능과 목적에 따라 정밀하게 제어할 수 있을 정도의 정확도(즉, 최적의 해)가 달성된 경우, 상기 강화학습을 종료(④)하고, 상기 강화학습을 종료한 상기 학습모델에 대한 학습 파라미터를 상기 연합강화학습 매니징 서버(300)로 리포팅한다.
- [0078] 또한 상기 연합강화학습 매니징 서버(300)는, 상기 강화학습을 종료한 디바이스 제어 장치(100)로부터 학습파라미터를 수신하면, 강화학습을 종료하지 않은 디바이스 제어 장치(100)로 전송(⑤)하는 학습파라미터 전이과정을 수행하여, 상기 강화학습을 종료하지 않은 디바이스 제어 장치(100)에서 상기 강화학습을 종료한 학습파라미터를 이용하여 상기 강화학습을 수행하여, 상기 정확도를 달성할 때까지 해당 강화학습을 지속할 수 있도록 한다.
- [0079] 상기 학습파라미터 전이과정을 통해 상기 강화학습을 종료하지 않은 디바이스 제어 장치(100)는, 자신의 강화학습 데이터뿐만 아니라, 상기 강화학습을 종료한 디바이스 제어 장치(100)에서 구성한 강화학습 데이터를 이용하여 상기 강화학습을 수행하는 효과가 있으며, 상기 정확도를 보다 빠르게 달성할 수 있다.
- [0080] 즉, 본 발명은, 상기 그레디언트 공유과정과, 상기 학습파라미터 전이과정을 통해, 복수의 디바이스 제어 장치(100)에서 개별적으로 수행되는 강화학습을 연합하여, 상기 강화학습에 대한 연합강화학습을 수행하도록 함으로써, 상기 복수의 디바이스 제어 장치(100)에서 수행되는 전체적인 강화학습을 조기에 달성하고, 상기 디바이스(300)를 정밀하게 제어할 수 있도록 하는 학습모델을 최종적으로 생성할 수 있도록 하는 것이다.
- [0081] 한편, 상기 그레디언트 공유과정과 상기 학습파라미터 전이과정은 순차적으로 수행되며, 복수회 수행될 수 있으며, 상기 강화학습이 종료된 디바이스 제어 장치(100)가 존재하는 경우에도, 상기 강화학습이 종료되지 않은 디바이스 제어 장치(100)에서, 상기 강화학습을 또 다시 연합한 연합강화학습을 수행할 수도 있다.
- [0082] 도 4는 본 발명의 일 실시예에 따른 연합강화학습을 통한 디바이스 제어 장치의 구성을 나타낸 블록도이다.
- [0083] 도 4에 도시한 바와 같이, 본 발명의 일 실시예에 따른 연합강화학습을 통한 디바이스 제어 장치(100)는, 상기 디바이스 제어 장치(100)와 연결된 디바이스(300)에게 제어정보를 전송하여 제어하는 디바이스 제어부(110), 상기 디바이스(300)로부터 상기 제어한 결과에 따른 상태정보를 수신하는 상태정보 수신부(120), 상기 수신한 상태정보에 따른 리워드를 생성하는 리워드 생성부(130), 상기 수신한 상태정보와, 상기 생성한 리워드를 토대로 상기 디바이스 제어를 위한 학습모델의 강화학습 데이터를 구성하는 강화학습 데이터 구성부(140), 디바이스 상태정보 제공부(150), 메모리(160) 및 상기 학습모델에 대한 강화학습과, 적어도 하나 이상의 또 다른 디바이스 제어 장치(100)에서 수행되는 강화학습을 연합하여 상기 강화학습에 대한 연합강화학습을 수행하여, 상기 디바이스를 정밀하게 제어하기 위한 학습모델을 최종적으로 생성하는 연합강화 학습부(170)를 포함하여 구성된다.
- [0084] 상기 디바이스 제어부(110)는, 상기 학습모델에 대한 강화학습을 수행하기 위해, 상기 디바이스(300)의 현재 상태정보를 상기 학습모델에 입력하여 상기 디바이스(300)를 제어하기 위한 제어정보를 추출하거나, 상기 연합강화 학습부(170)를 통해 최종적으로 생성된 학습모델을 이용하여, 상기 디바이스(300)를 제어하는 제어정보를 추출하여, 상기 디바이스(300)를 제어하는 기능을 수행한다.
- [0085] 또한 상기 상태정보 수신부(120)는, 상기 제어정보에 따라 현재 상태에서 다음 상태로 전이된 상기 디바이스(300)의 다음 상태정보를 수신하는 기능을 수행한다.
- [0086] 상기 복수의 디바이스 제어 장치(100)의 제어대상이 되는 각각의 디바이스(300)는, 특정 물건을 이송하는 무인 이송수단 등과 같이 유사한 특성을 가지고 동일한 목표를 가지는 장치를 의미하며, 상기 상태정보는, 상기 디바이스(300)의 위치정보, 상기 디바이스(300)의 구동을 위한 액추에이터의 각도, 각속도 등을 포함할 수 있음은

상술한 바와 같다.

- [0087] 또한 상기 리워드 생성부(140)는, 상기 제어정보를 토대로 전이된 상기 디바이스(300) 상태의 다음의 상태정보에 대한 리워드를 생성하는 기능을 수행한다.
- [0088] 상기 리워드는, 상기 수신한 디바이스(300)에 대한 다음 상태정보를 토대로 사전에 설정한 임계범위를 기준으로 상기 다음 상태정보에 대한 리워드를 계산함으로써, 생성된다.
- [0089] 또한 상기 리워드는, 상기 수신한 디바이스(300)의 다음 상태정보가 상기 임계범위 내에서 동작된 경우에는 사전에 설정한 플러스 보상값으로 생성되며, 상기 임계범위 외에서 동작된 경우에는 상기 임계범위의 근접한 정도에 따라 사전에 설정한 마이너스 보상값으로 차등적으로 생성된다.
- [0090] 또한 상기 강화학습 데이터 구성부(150)는, 상기 학습모델의 강화학습을 수행하기 위한 강화학습 데이터를 구성하는 기능을 수행한다.
- [0091] 상기 강화학습 데이터는, 상기 디바이스(300)의 현재 상태정보, 상기 현재 상태정보를 토대로 추출한 제어정보, 상기 제어정보를 토대로 제어된 상기 디바이스(300)로부터 수신한 다음 상태정보 및 상기 다음 상태정보에 따라 생성한 리워드를 포함한다.
- [0092] 또한 상기 강화학습 데이터 구성부(140)는, 상기 구성한 강화학습 데이터를 상기 연합강화 학습부(170)에게 제공하여, 상기 학습모델에 대한 강화학습을 지속적으로 수행할 수 있도록 한다.
- [0093] 또한 상기 연합강화 학습부(170)는, 상기 구성한 강화학습 데이터를 토대로 상기 학습모델에 대한 강화학습을 기본적으로 수행하며, 또 다른 적어도 하나 이상의 디바이스 제어 장치(100)에서 수행하는 강화학습과 연합한 연합강화학습을 수행함으로써, 상기 학습모델에 대한 강화학습을 조기해 완료하여 상기 디바이스(300)를 정밀하게 제어할 수 있도록 하는 학습모델을 최종적으로 생성하는 기능을 수행한다.
- [0094] 한편, 상기 연합강화학습은, 복수의 디바이스 제어 장치(100)에서 수행하는 강화학습에 대한 그레디언트를 공유하는 것과, 학습파라미터를 전이하는 것을 통해 수행되는 것으로, 도 3을 참조하여 설명하였으므로 더 이상의 상세한 설명은 생략하도록 한다.
- [0095] 또한 상기 연합강화 학습부(170)는, 강화학습 과정에서의 그레디언트를 계산하여, 연합강화학습 매니징 서버(200)로 리포팅하는 그레디언트 리포팅부(171), 상기 연합강화학습 매니징 서버(200)로부터 평균 그레디언트를 수신하는 평균 그레디언트 수신부(172), 상기 강화학습이 종료된 경우, 종료된 결과에 따른 학습파라미터를 리포팅하는 학습파라미터 리포팅부(173) 및 상기 연합강화학습 매니징 서버(300)로부터 학습파라미터를 수신하는 학습파라미터 수신부(174)를 포함하여 구성된다.
- [0096] 상기 그레디언트 리포팅부(171)는, 상기 연합강화 학습부(170)를 통해 상기 학습모델에 대한 강화학습을 수행하고 있는 도 중, 상기 연합강화학습 매니징 서버(200)에서 그레디언트 요청명령이 수신되는 경우, 상기 수행한 강화학습 과정에서 그레디언트를 계산하고, 상기 계산한 그레디언트를 상기 연합강화학습 매니징 서버(200)로 리포팅하는 기능을 수행한다.
- [0097] 또한 상기 평균 그레디언트 수신부(172)는, 상기 연합강화학습 매니징 서버(200)로부터 계산된 평균 그레디언트를 수신하는 기능을 수행한다.
- [0098] 여기서, 상기 평균 그레디언트는, 상기 복수의 디바이스 제어 장치(100)에서 리포팅되는 복수의 그레디언트에 대한 평균값을 의미하며, 상기 연합강화학습 매니징 서버(100)는, 상기 평균 그레디언트를 상기 복수의 디바이스 제어 장치(100)로 제공함으로써, 상기 평균 그레디언트를 복수의 디바이스 제어 장치(100)에서 공유할 수 있도록 한다.
- [0099] 또한 상기 연합강화 학습부(170)는, 상기 연합강화학습 매니징 서버(200)로부터 상기 평균 그레디언트가 수신되는 경우, 상기 수신한 평균 그레디언트를 사용하여 상기 강화학습을 지속하게 된다. 이는, 상기 복수의 디바이스 제어 장치(100)에서 각각 수행되는 전체 강화학습의 속도가 평균적인 속도로 진행되도록 한다.
- [0100] 또한 상기 학습파라미터 리포팅부(173)는, 상기 평균 그레디언트를 이용한 상기 강화학습이 종료된 경우, 상기 강화학습이 종료된 학습모델에 대한 학습파라미터를 상기 연합강화학습 매니징 서버(200)로 리포팅하는 기능을 수행한다.
- [0101] 또한 상기 학습파라미터 수신부(174)는, 상기 연합강화학습 매니징 서버(200)로부터 특정 디바이스 제어 장치

(100)에서 강화학습이 완료된 학습파라미터를 수신하는 기능을 수행한다.

- [0102] 여기서, 상기 특정 디바이스 제어 장치(100)에서 강화학습이 완료된 학습파라미터는, 상기 복수의 디바이스 제어 장치(100) 중 어느 하나에서 최초로 강화학습이 완료된 학습파라미터를 의미하며, 상기 연합강화 학습부(170)를 통해 상기 학습모델에 대한 강화학습이 종료되지 않은 경우에 수신된다.
- [0103] 즉, 상기 학습파라미터 수신부(174)를 통해 수신되는 학습파라미터는 상기 연합강화학습 매니징 서버(200)에 최초로 리포팅된 학습파라미터이며, 상기 연합강화학습 매니징 서버(200)는, 복수의 디바이스 제어 장치(100) 중 적어도 어느 하나에서, 기계학습이 종료된 학습파라미터를 수신하여, 상기 강화학습이 종료되지 않은 디바이스 제어 장치(100)로 상기 학습파라미터를 전송함으로써, 상기 학습파라미터를 전이하여, 상기 전이한 학습파라미터를 이용한 강화학습을 지속하도록 한다.
- [0104] 또한 상기 연합강화 학습부(170)는, 상기 평균 그레디언트를 이용한 강화학습이 종료되지 않은 상태에서, 상기 학습파라미터가 수신되는 경우, 상기 학습모델의 학습파라미터를 상기 수신한 학습파라미터로 설정하여, 상기 학습모델에 대한 강화학습을 계속하여 진행함으로써, 상기 디바이스(300)를 정밀하게 제어할 수 있도록 하는 학습모델을 최종적으로 생성하게 된다.
- [0105] 한편 상기 강화학습은, 강화학습에 최적화된 CNN(Convolutional Neural Network)기반의 DQN(Deep Q-Network)를 이용하여 수행되는 것이 바람직 하지만, ANN(Artificial Neural Network) 등과 같은 다양한 기계학습 네트워크를 통해 수행될 수 있다.
- [0106] 또한 상기 디바이스 제어부(110)는, 상기 생성한 학습모델을 토대로 상기 디바이스(300)의 기능이나 이용목적에 따라 상기 디바이스(300)를 정밀하게 제어하는 기능을 수행하며, 상기 디바이스 상태정보 제공부(150)는, 상기 제어한 디바이스(300)의 상태정보를 상기 연합강화학습 매니징 서버(200)로 전송하여, 상기 디바이스(300)를 실질적으로 제어한 상태를 모니터링할 수 있도록 한다.
- [0107] 이때, 상기 디바이스 상태정보는, 상기 디바이스(300)를 제어하기 전의 현재 상태정보, 상기 현재 상태정보를 토대로 추출한 제어정보 및 상기 제어정보에 따라 제어된 디바이스(300)의 다음 상태정보를 포함한다.
- [0108] 또한 상기 연합강화학습 매니징 서버(200)는, 상기 모니터링한 결과에 따라 상기 디바이스(300)를 정확하게 제어하지 못하는 경우, 상기 연합강화학습을 재수행할 수 있도록 하여, 상기 디바이스 제어 장치(100)가 상기 디바이스(300)의 에이징이나, 디바이스(300)의 주변 환경에 적응하여 상기 디바이스(300)를 정밀하게 제어할 수 있도록 한다.
- [0109] 또한 상기 메모리(160)는, 상기 학습모델과, 강화학습 데이터를 저장하고, 상기 디바이스 제어 장치(100)의 동작을 위한 관련 정보들을 저장하는 기능을 수행한다.
- [0110] 도 5는 본 발명의 일 실시예에 따른 연합강화학습 매니징 서버의 구성을 나타낸 블록도이다.
- [0111] 도 5에 도시한 바와 같이, 본 발명의 일 실시예에 따른 연합강화학습 매니징 서버(200)는, 복수의 디바이스 제어 장치(100)로부터 리포팅된 그레디언트를 수신하는 그레디언트 수신부(210), 상기 수신한 그레디언트를 평균하여 상기 복수의 디바이스 제어 장치(100)로 공유하는 그레디언트 공유부(220), 상기 복수의 디바이스 제어 장치(100)로부터 리포팅된 학습파라미터를 수신하는 학습파라미터 수신부(230), 상기 수신한 학습파라미터를 상기 강화학습이 종료되지 않은 디바이스 제어 장치(100)로 제공하는 학습파라미터 제공부(240), 상기 강화학습결과로 생성한 학습모델을 통해 상기 디바이스 제어 장치(100)별로 디바이스(300)를 제어한 결과인 디바이스 상태정보를 수신하는 디바이스 상태정보 수신부(250), 상기 수신한 디바이스 상태정보를 토대로 상기 디바이스(300)에 대한 제어상태를 모니터링하는 디바이스 상태정보 모니터링부(260)를 포함하여 구성된다.
- [0112] 또한 상기 그레디언트 수신부(210)는, 상기 강화학습을 수행하는 복수의 디바이스 제어 장치(100)로 상기 강화학습에서 계산되는 그레디언트를 요청하여 수신하는 기능을 수행한다.
- [0113] 또한 상기 그레디언트 공유부(220)는, 상기 복수의 디바이스 제어 장치(100)로부터 수신되는 복수의 그레디언트에 대한 평균을 계산하여, 평균 그레디언트를 산출한다.
- [0114] 또한 상기 그레디언트 공유부(220)는, 상기 산출한 평균 그레디언트를 상기 복수의 디바이스 제어 장치(100)로 제공함으로써, 그레디언트를 복수의 디바이스 제어 장치(100)로 공유하는 과정을 수행한다.
- [0115] 즉, 상기 그레디언트를 공유하는 그레디언트 공유과정은, 상기 복수의 그레디언트를 평균한 평균 그레디언트를 상기 복수의 디바이스 제어 장치(100)로 제공함으로써, 수행되는 것이다.

- [0116] 또한 상기 학습파라미터 수신부(230)는, 상기 공유한 그레디언트를 이용하여 상기 강화학습을 지속 진행한 복수의 디바이스 제어 장치(100)에서 상기 강화학습이 종료된 적어도 하나의 디바이스 제어 장치(100)로부터 상기 강화학습 종료에 따른 학습파라미터를 수신하는 기능을 수행한다.
- [0117] 또한 상기 학습파라미터 제공부(240)는, 상기 수신한 학습파라미터를 상기 강화학습이 종료되지 않은 적어도 하나 이상의 디바이스 제어 장치(100)로 제공함으로써, 상기 수신한 학습파라미터를 상기 적어도 하나 이상의 디바이스 제어 장치(100)로 전이시키는 학습파라미터 전이과정을 수행한다.
- [0118] 즉, 상기 연합강화학습 매니징 서버(200)는, 상기 그레디언트 공유과정과, 상기 학습파라미터 전이과정을 통해 상기 디바이스 제어 장치(100)별로 수행되는 강화학습을 연합하여, 상기 디바이스 제어 장치(100)에서 상기 강화학습에 대한 연합강화학습을 수행하도록 함으로서, 상기 디바이스 제어 장치(100)별로 수행되는 전체적인 강화학습을 정확하고 조기에 완료할 수 있도록 하여, 실제 디바이스(300)를 정밀하게 제어할 수 있도록 한다.
- [0119] 또한 상기 디바이스 상태정보 수신부(250)는, 상기 연합강화학습을 수행한 결과에 따라 디바이스 제어 장치(100)별로 생성한 학습모델을 토대로, 상기 디바이스(300)를 각각 제어한 디바이스 상태정보를 수신하는 기능을 수행한다.
- [0120] 또한 상기 디바이스 상태정보 모니터링부(260)는, 상기 수신한 디바이스 상태정보를 토대로 상기 디바이스(300)를 제어한 제어상태를 모니터링하는 기능을 수행하며, 상기 모니터링한 결과에 따라 상기 디바이스(300)가 정확하게 제어되지 않은 경우, 상기 연합강화학습을 재수행하도록 하는 기능을 수행한다.
- [0121] 이때, 상기 디바이스 상태정보 모니터링부(260)는, 상기 디바이스 상태정보를 모니터링하여, 상기 제어정보에 따라 상기 디바이스(300)의 현재 상태정보에서 다음 상태정보로 전이될 때, 사전에 설정한 임계범위에 따라 전이되었는지에 대한 여부를 판단함으로써, 상기 연합강화학습에 대한 재수행 여부를 판단하게 된다.
- [0122] 즉, 상기 디바이스 상태정보 모니터링부(260)는, 상기 다음 상태정보가 상기 임계범위를 초과하는 경우에는 상기 연합강화학습을 재수행할 수 있도록 하여, 상기 디바이스(300)의 이용에 따른 에이징이나, 주변 환경 변화에 적응하여, 상기 디바이스(300)를 지속적으로 정밀하게 제어할 수 있도록 하는 것이다.
- [0123] 도 6은 본 발명의 일 실시예에 따른 연합강화학습을 통한 다중 디바이스 제어하는 절차를 나타낸 흐름도이다.
- [0124] 도 6에 도시한 바와 같이, 본 발명의 일 실시예에 따른 연합강화학습을 통한 다중 디바이스를 제어하는 절차는 우선, 복수의 디바이스 제어 장치(100)에서, 상기 각 디바이스 제어 장치(100)와 연결된 디바이스를 각각 제어하기 위한 강화학습을 수행하는 단계를 수행한다(S110).
- [0125] 상기 강화학습을 수행하는 단계는, 사전에 준비된 각각의 학습모델에 대해 수행되는 것이며, 상기 강화학습을 수행하는 과정에 대해서는 도 2를 참조하여 설명하였으므로 더 이상의 상세한 설명은 생략하도록 한다.
- [0126] 다음으로, 상기 디바이스 제어 장치(100)는, 연합강화학습 매니징 서버(200)로부터 그레디언트 요청이 있는 경우, 상기 강화학습을 수행하는 강화학습 과정에서의 그레디언트를 계산하여 상기 연합강화학습 매니징 서버(200)로 리포팅하는 단계를 수행한다(S120).
- [0127] 상기 그레디언트는, 상기 각 디바이스 제어 장치(100)별 강화학습이 수행되는 속도를 의미한다.
- [0128] 다음으로, 상기 연합강화학습 매니징 서버(200)에서, 상기 리포팅된 복수의 그레디언트를 평균을 계산한 평균 그레디언트를 산출하고, 상기 산출한 평균 그레디언트를 상기 복수의 디바이스 제어 장치(100)로 제공함으로써, 상기 평균 그레디언트를 공유하는 단계를 수행한다(S130).
- [0129] 즉, 상기 연합강화학습 매니징 서버(200)는, 상기 평균 그레디언트를 공유하는 그레디언트 공유과정을 수행하여, 상기 각 디바이스 제어 장치(100)에서 수행되는 강화학습이 평균적인 속도로 진행되도록 한다.
- [0130] 다음으로, 상기 복수의 디바이스 제어 장치(100)에서, 상기 연합강화학습 매니징 서버(200)를 통해 공유되는 상기 평균 그레디언트를 이용하여 상기 강화학습을 지속하고, 상기 강화학습이 종료된 경우, 상기 종료된 결과에 따른 학습파라미터를 상기 연합강화학습 매니징 서버(200)로 리포팅하는 단계를 수행한다(S140).
- [0131] 즉, 상기 복수의 디바이스 제어 장치(100)는, 상기 공유된 평균 그레디언트를 이용하여 상기 강화학습을 지속하며, 상기 지속한 결과 상기 강화학습의 대상이 되는 학습모델이 상기 디바이스(300)를 정밀하게 제어하기 위한 최적의 상태(즉, 상기 디바이스를 정밀하게 제어할 정도의 정확도)가 달성되면, 상기 강화학습을 종료하여, 상기 디바이스(300)를 정밀하게 제어하기 위한 학습모델을 최종적으로 생성하게 되며, 상기 강화학습을 종료한 학

답모델의 학습파라미터를 상기 연합강화학습 매니징 서버(200)로 전송하는 것이다.

- [0132] 다음으로, 상기 연합강화학습 매니징 서버(200)에서 상기 강화학습이 종료되지 않은 디바이스 제어 장치(100)로 상기 수신한 학습파라미터를 전송하는 단계를 수행한다(S150).
- [0133] 즉, 상기 연합강화학습 매니징 서버(200)는, 상기 평균 그래디언트를 이용하여 강화학습을 조기에 종료한 적어도 하나의 디바이스 제어 장치(100)로부터 학습파라미터를 수신하고, 상기 수신한 학습파라미터를 상기 강화학습을 종료하지 않은 적어도 하나 이상의 디바이스 제어 장치(100)로 전이시키는 학습파라미터 전이과정을 수행하는 것이다.
- [0134] 다음으로, 상기 복수의 디바이스 제어 장치(100)에서, 상기 연합강화학습 매니징 서버(200)로부터 전송되어 수신되는 상기 학습파라미터를 이용하여 상기 강화학습을 지속하여, 상기 학습모델에 대한 상기 정확도가 달성되면 해당 강화학습을 종료하게 된다.
- [0135] 상기에서 설명한 것과 같이, 본 발명은, 상기 S110 단계 내지 S160 단계를 통해 복수의 디바이스 제어 장치(100)에서 개별적으로 수행되는 강화학습을 연합하여, 상기 강화학습에 대한 연합강화학습을 수행하도록 함으로써, 상기 복수의 디바이스 제어 장치(100)에서 수행되는 복수의 강화학습을 조기에 종료시킴과 동시에, 상기 디바이스(100) 제어를 위한 높은 정확도를 가지는 학습모델을 각각 생성할 수 있도록 한다.
- [0136] 다음으로, 상기 복수의 디바이스 제어 장치(100)에서, 상기 연합강화학습을 통해 강화학습이 진행되어 종료된 결과에 따라, 최종적으로 생성되는 학습모델을 이용하여 상기 디바이스(300)를 정밀하게 제어하는 단계를 수행한다(S170).
- [0137] 상기 디바이스(300)를 제어하는 것은, 상기 디바이스(300)의 기능이나 이용목적에 따라 상기 생성한 학습모델에 상기 디바이스(300)의 현재 상태정보를 입력하여, 상기 학습모델을 통해 상기 현재 상태정보에 대한 제어정보를 추출하고, 상기 추출한 제어정보에 따라 상기 디바이스(300)를 제어하는 과정을 반복함으로써, 수행된다.
- [0138] 이상에서 설명한 바와 같이, 본 발명의 일 실시예에 따른 연합강화학습을 통한 다중 디바이스 제어 시스템 및 그 방법은, 복수의 디바이스 제어 장치에서 디바이스를 제어하기 위한 학습모델에 대한 강화학습을 수행할 때, 상기 복수의 디바이스 제어 장치별로 수행되는 강화학습을 연합하여, 상기 강화학습에 대한 연합강화학습을 수행하도록 함으로써, 상기 복수의 디바이스 제어 장치에서 각각 수행되는 복수의 강화학습을 조기에 완료하여 상기 디바이스를 정밀하게 제어할 수 있도록 하는 학습모델을 생성할 수 있도록 하는 효과가 있다.
- [0139] 상기에서는 본 발명에 따른 바람직한 실시예를 위주로 상술하였으나, 본 발명의 기술적 사상은 이에 한정되는 것은 아니며 본 발명의 각 구성요소는 동일한 목적 및 효과의 달성을 위하여 본 발명의 기술적 범위 내에서 변경 또는 수정될 수 있을 것이다.
- [0140] 또한, 이상에서는 본 발명의 바람직한 실시예에 대하여 도시하고 설명하였지만, 본 발명은 상술한 특징의 실시예에 한정되지 아니하며, 청구범위에서 청구하는 본 발명의 요지를 벗어남이 없이 당해 발명이 속하는 기술분야에서 통상의 지식을 가진 자에 의해 다양한 변형 실시가 가능한 것은 물론이고, 이러한 변형 실시들은 본 발명의 기술적 사상이나 전망으로부터 개별적으로 이해되어서는 안 될 것이다.

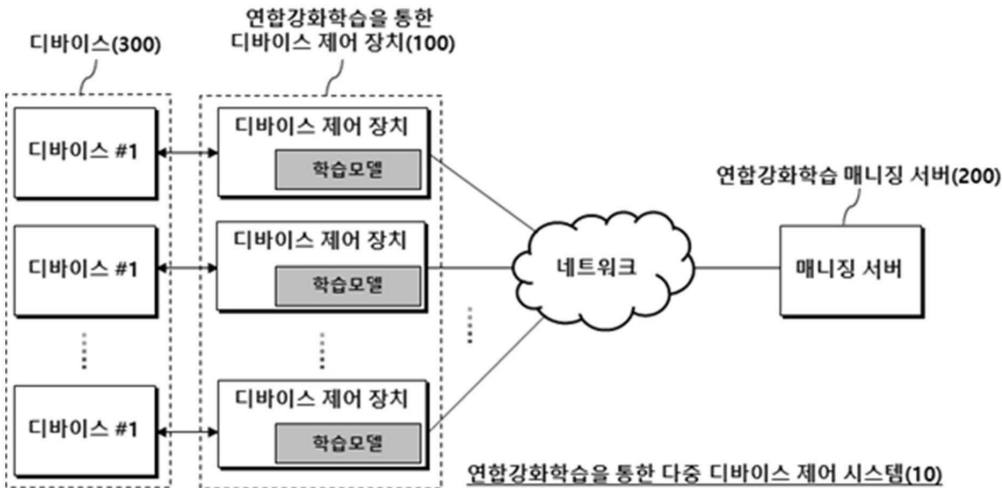
부호의 설명

- [0141] 10: 연합강화학습을 통한 다중 디바이스 제어 시스템
- 100: 연합강화학습을 통한 디바이스 제어 장치 110: 디바이스 제어부
- 120: 상태정보 수신부 130: 리워드 생성부
- 140: 강화학습 데이터 구성부 150: 디바이스 상태정보 제공부
- 160: 메모리 170: 연합강화 학습부
- 171: 그래디언트 리포팅부 172: 평균 그래디언트 수신부
- 173: 학습파라미터 리포팅부 174: 학습파라미터 수신부
- 200: 연합강화학습 매니징 서버 210: 그래디언트 수신부
- 220: 그래디언트 공유부 230: 학습파라미터 수신부

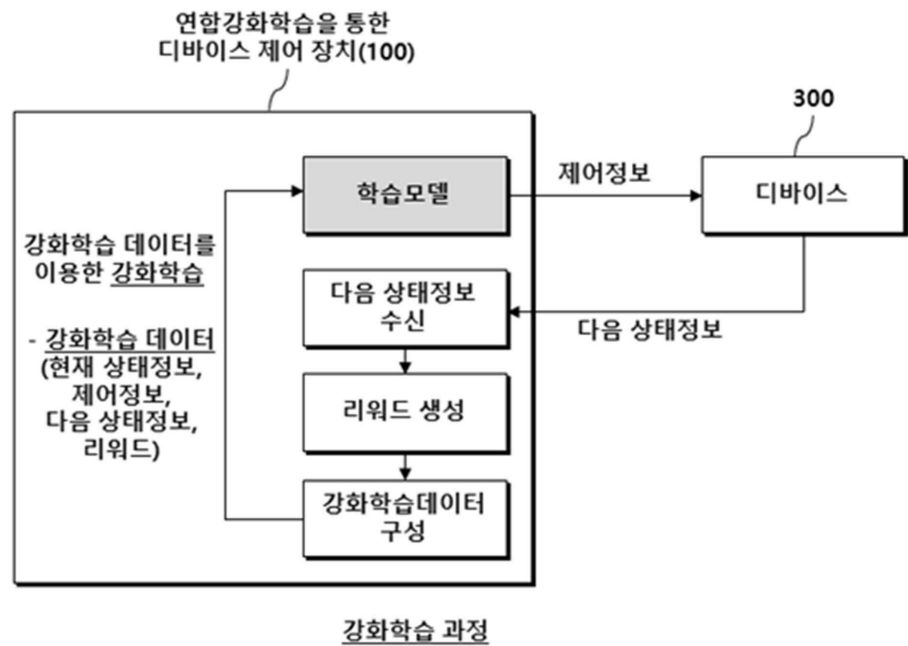
- 240: 학습파라미터 제공부
- 250: 디바이스 상태정보 수신부
- 260: 디바이스 상태정보 모니터링부

도면

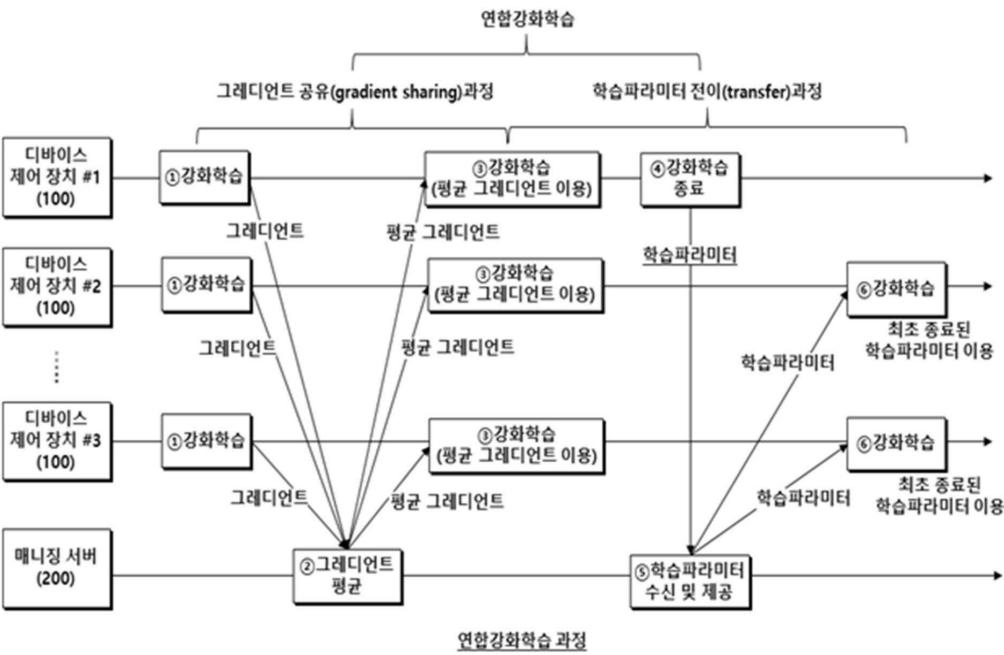
도면1



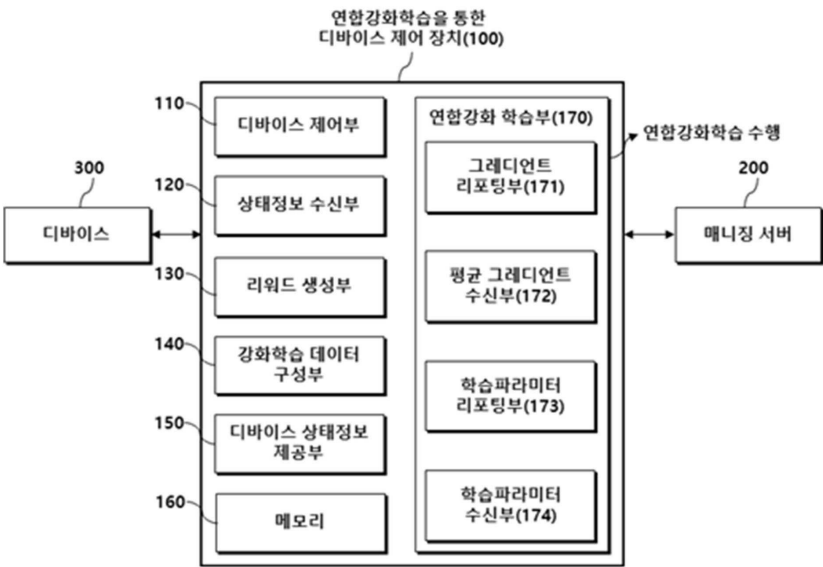
도면2



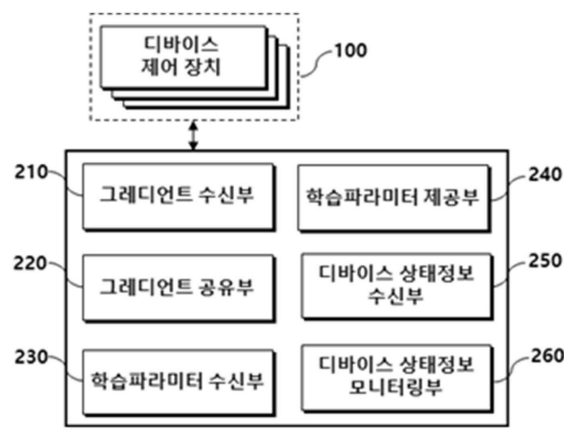
도면3



도면4



도면5



도면6

