(19) 대한민국특허청(KR)
(12) 등록특허공보(B1)

(45) 공고일자 2023년03월27일
(11) 등록번호 10-2515139
(24) 등록일자 2023년03월23일

(51) 국제특허분류(Int. Cl.)
G06N 20/00 (2019.01) *G06N 3/00* (2022.01)
G06N 3/08 (2023.01)
(52) CPC특허분류
G06N 20/00 (2021.08)
G06N 3/006 (2023.01)
(21) 출원번호 10-2022-0112155
(22) 출원일자 2022년09월05일
심사청구일자 2022년09월05일
(56) 선행기술조사문헌
KR1020010047761 A*
"전장 디지털트윈을 활용한 지휘결심지원기술 개발 방안", 한국통신학회 학술대회논문집, 2021.*
KR1020220117123 A
KR102365169 B1
*는 심사관에 의하여 인용된 문헌

(73) 특허권자
세종대학교산학협력단
서울특별시 광진구 능동로 209 (군자동, 세종대학교)
(72) 발명자
이현석
서울특별시 성동구 상원길 63, 107동 602호(성수동1가, 쌍용아파트)
김지완
경기도 용인시 기흥구 동백죽전대로527번길 81, 103동 1102호(중동, 신동백 서해그랑블 1차)
(뒷면에 계속)
(74) 대리인
민영준

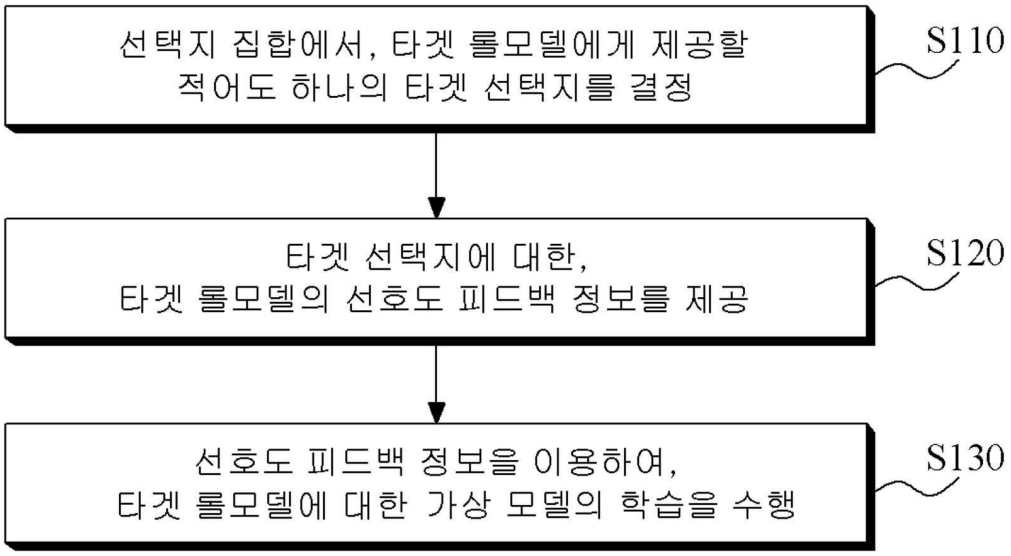
전체 청구항 수 : 총 13 항 심사관 : 박승철

(54) 발명의 명칭 **강화 학습 기반의 가상 모델 학습 방법 및 가상 모델 서비스 제공 방법**

(57) 요 약

사용자의 타겟 롤모델에 대한 가상 모델을 학습하는 방법 및 가상 모델 이용한 서비스 제공 방법이 개시된다. 강화 학습 기반의 가상 모델 학습 방법은 선택지 집합에서, 타겟 롤모델에게 제공할 적어도 하나의 타겟 선택지를 결정하는 단계; 상기 타겟 선택지에 대한, 상기 타겟 롤모델의 선호도 피드백 정보를 제공받는 단계; 및 상기 선호도 피드백 정보을 이용하여, 상기 타겟 롤모델에 대한 가상 모델의 학습을 수행하는 단계를 포함하며, 상기 가상 모델은, 입력 데이터에 대한 선호도값을 출력하는 강화 학습 기반의 선호도 네트워크를 포함한다.

대 표 도 - 도1



(52) CPC특허분류
G06N 3/08 (2023.01)
(72) 발명자
이승진
서울특별시 광진구 광나루로14길 15, 502호(화양동, YN웰하우스)

제갈홍
경기도 안양시 만안구 안양천서로 177, 212동 1004호(안양동, 래미안 안양 메가트리아)

이 발명을 지원한 국가연구개발사업

과제고유번호	1711152732
과제번호	2021-0-01816-002
부처명	과학기술정보통신부
과제관리(전문)기관명	정보통신기획평가원
연구사업명	정보통신방송혁신인재양성
연구과제명	메타버스 자율트윈 핵심기술 연구
기 여 율	1/1
과제수행기관명	세종대학교 산학협력단
연구기간	2022.01.01 ~ 2022.12.31

공지예외적용 : 있음

명 세 서

청구범위

청구항 1

컴퓨팅 장치에 의해 수행되는, 강화 학습 기반의 가상 모델 학습 방법에 있어서,

선택지 집합에서, 타겟 롤모델에게 제공할 적어도 하나의 타겟 선택지를 결정하는 단계;

상기 타겟 선택지에 대한, 상기 타겟 롤모델의 선호도 피드백 정보를 제공받는 단계; 및

상기 선호도 피드백 정보을 이용하여, 상기 타겟 롤모델에 대한 가상 모델의 학습을 수행하는 단계를 포함하며,

상기 타겟 선택지를 결정하는 단계는

상기 타겟 롤모델에 대한 특징 정보, 상기 타겟 롤모델에 의해 수행된 의사 결정에 대한 특징 정보 및 상기 타겟 롤모델의 의사 결정 시점에 대한 컨텍스트 정보를 이용하여, 상기 타겟 선택지를 결정하며,

상기 가상 모델은,

입력 데이터에 대한 선호도값을 출력하는 강화 학습 기반의 선호도 네트워크를 포함하는

강화 학습 기반의 가상 모델 학습 방법.

청구항 2

삭제

청구항 3

제 1항에 있어서,

상기 타겟 선택지를 결정하는 단계는

상기 선택지 집합에 포함된 선택지, 상기 타겟 롤모델에 대한 특징 정보, 상기 의사 결정에 대한 특징 정보 및 상기 컨텍스트 정보를 상기 선호도 네트워크에 입력하여, 상기 선택지 집합에 포함된 선택지 각각에 대한 선호도값을 획득하는 단계; 및

상기 선호도값에 따라서, 상기 타겟 선택지를 결정하는 단계

를 포함하는 강화 학습 기반의 가상 모델 학습 방법.

청구항 4

제 1항에 있어서,

상기 타겟 롤모델에 대한 특징 정보는

상기 타겟 롤모델의 성별, 나이 및 국적 중 적어도 하나를 포함하는

강화 학습 기반의 가상 모델 학습 방법.

청구항 5

제 1항에 있어서,

상기 타겟 롤모델에 의해 수행된 의사 결정에 대한 특징 정보는

상기 타겟 롤모델의 의사 결정에 의해 선택된 뉴스의 카테고리, 보도 시간, 키워드 및 언론사 중 적어도 하나를 포함하는

강화 학습 기반의 가상 모델 학습 방법.

청구항 6

제 1항에 있어서,

상기 컨텍스트 정보는

상기 의사 결정 시점의 날짜, 시간 및 날씨 중 적어도 하나를 포함하는

강화 학습 기반의 가상 모델 학습 방법.

청구항 7

제 1항에 있어서,

상기 선호도 피드백 정보는

상기 타겟 선택지 중에서, 상기 타겟 롤모델에 의해 선택된 선택지에 대해 제1보상값이 할당되고, 상기 타겟 롤모델에 의해 선택되지 않은 선택지에 대해 제2보상값이 할당되는 정보인

강화 학습 기반의 가상 모델 학습 방법.

청구항 8

제 7항에 있어서,

상기 가상 모델의 학습을 수행하는 단계는

상기 타겟 선택지, 상기 타겟 롤모델에 대한 특징 정보, 상기 의사 결정에 대한 특징 정보 및 상기 컨텍스트 정보를 상기 선호도 네트워크에 입력하여, 상기 타겟 선택지에 대한 선호도값을 획득하는 단계; 및

상기 선호도 피드백 정보 및 상기 선호도값으로부터 계산된 손실값이 최소가 되도록, 상기 선호도 네트워크에 대한 학습을 수행하는 단계

를 포함하는 강화 학습 기반의 가상 모델 학습 방법.

청구항 9

제 1항에 있어서,

상기 가상 모델은

상기 타겟 롤모델에 대한 디지털 트윈인

강화 학습 기반의 가상 모델 학습 방법.

청구항 10

컴퓨팅 장치에 의해 수행되는, 강화 학습 기반의 가상 모델 서비스 제공 방법에 있어서,

사용자의 요청에 따라, 저장 장치에 저장된 타겟 롤모델에 대한 특징 정보, 상기 타겟 롤모델에 의해 수행된 의사 결정에 대한 특징 정보 및 상기 사용자의 선호도값 요청 시점에 대한 컨텍스트 정보를 로딩하는 단계; 및

타겟 선택지 및 상기 로딩된 정보를 상기 타겟 롤모델에 대한 가상 모델에 입력하여, 상기 타겟 선택지에 대한 상기 가상 모델의 선호도값을 생성하는 단계를 포함하며,

상기 가상 모델은, 상기 타겟 선택지에 대한 선호도값을 제공하는, 강화 학습 기반의 선호도 네트워크를 포함하며,

상기 컨텍스트 정보는

선호도값 의사 결정 시점의 날짜, 시간 및 날씨 중 적어도 하나를 포함하는

강화 학습 기반의 가상 모델 서비스 제공 방법.

청구항 11

제 10항에 있어서,

상기 선호도 네트워크는

상기 타겟 선택지 및 상기 로딩된 정보를 입력받는 입력층;

은닉층; 및

상기 타겟 선택지에 대한 선호도값을 출력하는 출력층

을 포함하는 강화 학습 기반의 가상 모델 서비스 제공 방법.

청구항 12

제 10항에 있어서,

상기 타겟 선택지 중에서, 상기 선호도값에 따라 선택된 선택지를 상기 사용자에게 제공하는 단계

를 더 포함하는 강화 학습 기반의 가상 모델 서비스 제공 방법.

청구항 13

제 10항에 있어서,

상기 의사 결정에 대한 특징 정보는

미리 설정된 카테고리의 선택지 중에서, 상기 타겟 롤모델의 의사 결정에 의해 선택된 선택지에 대한 특징 정보를 포함하는

강화 학습 기반의 가상 모델 서비스 제공 방법.

청구항 14

제 13항에 있어서,

상기 의사 결정에 대한 특징 정보는

상기 타겟 롤모델의 의사 결정에 의해 선택된 뉴스의 카테고리, 보도 시간, 키워드 및 언론사 중 적어도 하나를 포함하는

강화 학습 기반의 가상 모델 서비스 제공 방법.

발명의 설명

기 술 분 야

[0001] 본 발명은 사용자의 타겟 롤모델에 대한 가상 모델을 학습하는 방법 및 가상 모델을 이용한 서비스 제공 방법에 관한 것이다.

배 경 기 술

[0003] 메타버스(metaverse)란 가공·초월을 의미하는 메타(Meta)와 세계를 의미하는 유니버스(Universe)의 합성어로서, 가상과 현실이 융복합 된 디지털 세계, 초월 세계를 의미한다. 그리고 최근 메타버스와 함께 디지털 트윈(digital twin) 또한 주목받고 있다.

[0004] 디지털 트윈이란 현실 세계에서 실존하는 사람, 자동차, 제조설비 등을 디지털 공간에 실물과 똑같은 쌍둥이를 만들고, 현실 객체의 동작과 행위를 가상 세계에서도 실현시켜서, 현실 세계와 가상 세계의 쌍둥이 개체가 서로의 변화를 동기화시키는 것이다. 현재 디지털 트윈 4차 산업혁명을 견인하는 IOT, 인공지능과 같은 기술이 발전함에 따라, 이들을 응용하여 산업 현장에서 생산성, 경제성, 안정성을 향상하고자 하는 요구를 충족하기 위한 기술 트렌드로서 디지털 트윈 기술이 주목받고 있다.

[0005] 디지털 트윈을 구현하기 위해, 모방할 대상이나 시스템의 기초가 되는 물리학을 연구하고 연구 데이터를 사용해 디지털 공간에서 실제 원본을 시뮬레이션하는 수학적 모델을 개발한다. 따라서 관심대상 물리적 객체의 데이터와 행위를 추상화한 디지털 모델을 만들고 시뮬레이션을 통해 해당 실 체계의 운영 관련 예측/최적화를 달성하고자 한다는 점이, 종래의 시뮬레이션 모델에 대응되어, 디지털 트윈은 주로 스마트 시티 사전 검증, 제조 부문 제조 공정 효율성 제고 및 최적화, 기계 고장 진단 등 산업 전반에 걸쳐 활용되고 있다.

[0006] 디지털 트윈 구현에 필요한 디지털 모델은 일반적으로 인공지능(머신러닝) 데이터 기반의 모델링을 통해 만들어진다. 이러한 데이터 모델링 기술은 대상 시스템에서 관찰된 데이터 내의 특정 패턴 혹은 데이터 요소들간의 상관관계 분석을 통해 진단 및 예측 분석이 가능하다.

[0007] 최근에는 대상 객체로부터 수집되는 빅데이터를 머신러닝 알고리즘으로 학습하여 입력 데이터와 출력 데이터 간의 비선형적이고 복잡한 상관관계를 멀티 계층의 인공신경망 형태로 모델링하여, 분석 및 예측에 활용하는 사례가 증가하고 있다. 머신 러닝을 통해 시뮬레이션을 튜닝하고 축소하며 온라인 실시간 시뮬레이션을 가능하게 하며, 관찰 데이터의 학습을 통해 모델링 대상에 대한 불충분한 지식을 보완하는 방식으로 시뮬레이션 모델을 완성할 수 있기 때문이다. 하지만 이렇게 디지털 트윈의 핵심이 목적에 맞는 모델을 만드는 것으로, 모델링 시뮬레이션 및 수학적 모델 개발, 데이터 모델링이 중요 하지만 국내의 경우 주로 외국에서 도입된 시뮬레이션 모델을 사용하여 참조할 수 있는 모델이 없는 신규 모델 개발에 어려움을 겪고 있고 모델 개발보다는 활용 능력에 치중하고 있다.

[0008] 또한, 위와 같이 시각화가 중요한 제조 산업 분야로부터 적용되면서 발전되고 있기 때문에 디지털 트윈 응용에서는 가상 모델 시각화 해주는 기능이 주로 강조되고 있다. 디지털 트윈이 물리적 대상의 설계, 제조/생산, 운영 및 유지 보수 등 전체 수명주기를 지원하는 기술로 활용됨에 따라 3D 시각화 모델 생성 기능이 강조되고 있고 가상현실, 증강현실 등과 같은 가시화 기능의 집중하여 발전하고 있다.

[0009] 즉, 사람에 대한 디지털 트윈 구현시 실제 사람의 가치판단과 같은 선호도 및 생각, 예측, 판단을 고려하는 모델 구현보다는, 대부분 대상을 그대로 복원하는 시각화 기술에 치중한 신체적인 트윈에 기술 개발이 집중되고 있다.

[0010] 관련 선행문헌으로 대한민국 등록특허 제10-2390615호, 제10-2329074호, 대한민국 공개특허 제2022-0100226호, 제2020-0094758호가 있다.

발명의 내용

해결하려는 과제

[0012] 본 발명은 롤모델의 외형이 아닌 내적 특성이 반영되도록, 롤모델에 대한 가상 모델을 학습하는 방법을 제공하기 위한 것이다.

[0013] 또한 본 발명은 사용자가 장소와 시간에 구애됨이 없이, 롤모델에 대한 정보를 획득할 수 있도록, 롤모델에 대한 가상 모델을 사용자에게 서비스하는 방법을 제공하기 위한 것이다.

과제의 해결 수단

[0015] 상기한 목적을 달성하기 위한 본 발명의 일 실시예에 따르면, 선택지 집합에서, 타겟 롤모델에게 제공할 적어도 하나의 타겟 선택지를 결정하는 단계; 상기 타겟 선택지에 대한, 상기 타겟 롤모델의 선호도 피드백 정보를 제공받는 단계; 및 상기 선호도 피드백 정보을 이용하여, 상기 타겟 롤모델에 대한 가상 모델의 학습을 수행하는 단계를 포함하며, 상기 가상 모델은, 입력 데이터에 대한 선호도값을 출력하는 강화 학습 기반의 선호도 네트워크를 포함하는 강화 학습 기반의 가상 모델 학습 방법이 제공된다.

[0016] 또한 상기한 목적을 달성하기 위한 본 발명의 다른 실시예에 따르면, 사용자의 요청에 따라, 저장 장치에 저장된 타겟 롤모델에 대한 특징 정보, 상기 타겟 롤모델에 의해 수행된 의사 결정에 대한 특징 정보 및 상기 사용자의 요청 시점에 대한 컨텍스트 정보를 로딩하는 단계; 및 타겟 선택지 및 상기 로딩된 정보를 상기 타겟 롤모델에 대한 가상 모델에 입력하여, 상기 타겟 선택지에 대한 상기 가상 모델의 선호도값을 생성하는 단계를 포함하며, 상기 가상 모델은, 상기 타겟 선택지에 대한 선호도값을 제공하는, 강화 학습 기반의 선호도 네트워크를 포함하는 강화 학습 기반의 가상 모델 서비스 제공 방법이 제공된다.

발명의 효과

[0018] 본 발명의 일실시예에 따르면, 롤모델의 내적 특성이 반영된 롤모델에 대한 가상 모델이 제공될 수 있다.

[0019] 또한 본 발명의 일실시예에 따르면, 롤모델에 대한 가상 모델을 사용자에게 제공함으로써, 사용자의 의사 결정에 도움이 될 수 있는 롤모델에 대한 정보를 사용자가 시간과 장소의 구애됨이 없이 획득할 수 있다.

도면의 간단한 설명

[0021] 도 1은 본 발명의 일실시예에 따른 강화 학습 기반의 가상 모델 학습 방법을 설명하기 위한 도면이다.

도 2는 본 발명의 일실시예에 따른 강화 학습 모델을 설명하기 위한 도면이다.

도 3은 본 발명의 일실시예에 따른 가상 모델 학습 방법의 의사 코드를 나타내는 도면이다.

도 4는 본 발명의 일실시예에 따른 강화 학습 기반의 가상 모델 서비스 제공 방법을 설명하기 위한 도면이다.

도 5는 본 발명의 일실시예에 따른 가상 모델이 챗봇 형태로 서비스되는 일실시예를 나타내는 도면이다.

발명을 실시하기 위한 구체적인 내용

[0022] 본 발명은 다양한 변경을 가할 수 있고 여러 가지 실시예를 가질 수 있는 바, 특정 실시예들을 도면에 예시하고 상세한 설명에 상세하게 설명하고자 한다. 그러나, 이는 본 발명을 특정한 실시 형태에 대해 한정하려는 것이 아니며, 본 발명의 사상 및 기술 범위에 포함되는 모든 변경, 균등물 내지 대체물을 포함하는 것으로 이해되어야 한다. 각 도면을 설명하면서 유사한 참조부호를 유사한 구성요소에 대해 사용하였다.

[0024] 과거로부터, 사람들은 닮고자 하는 인물인 롤모델의 행동을 분석하고, 행동 그대로를 따라하는 등 자신의 삶을 발전시키기 위해, 타인의 삶의 방식을 배우고 모사하는 경향이 있다. 하지만, 사람들이 현실 세계에서 롤모델을 실제로 만나기도 어려울 뿐만 아니라, 실제로 롤모델을 만나 그들의 삶에 대해 배우기에는 환경, 시간적 제약이 따른다.

[0025] 이에 본 발명은 시간과 장소에 구애됨이 없이, 사용자가 롤모델로부터 롤모델의 성향이나 특성에 대한 정보를 획득할 수 있는 방법을 제안하며, 본 발명의 일실시예는 롤모델의 성향이나 특성으로서, 롤모델의 선호도를 사용자가 용이하게 획득할 수 있는 방법을 제안한다. 사용자가 획득한 롤모델의 선호도는 사용자의 의사 결정에 활용될 수 있다.

[0026] 본 발명의 일실시예는 사용자가 시간과 장소에 구애됨이 없이, 롤모델과 접촉할 수 있도록 가상 세계에서 구현되는 롤모델에 대한 가상 모델을 이용한다. 이러한 가상 모델은 일실시예로서, 타겟 롤모델에 대한 디지털 트윈이나 챗봇 등의 형태로 구현될 수 있다.

[0027] 본 발명의 일실시예에서 롤모델에 대한 가상 모델은, 강화 학습 모델을 이용해 롤모델의 선호도 정보를 학습한다. 일반적으로 사람의 특성이나 특징이, 지금까지의 선택과 행동을 통해 유추될 수 있다는 점에 착안하여, 본 발명의 일실시예는 롤모델의 과거의 의사 결정에 대한 정보를 이용해, 가상 모델에 대한 학습을 수행한다.

[0028] 학습된 가상 모델은 사용자의 의사 결정이 필요한 타겟 선택지에 대한 선호도를 제공한다. 예컨대, 사용자가 뉴

스를 선택하여 읽고 싶은 상황에서, 타겟 선택지로 다양한 뉴스들을 가상 모델에 제공하면, 가상 모델은 다양한 뉴스들 각각에 대한 선호도를 사용자에게 제공할 수 있다. 사용자는 다양한 뉴스들 중에서 가상 모델의 선호도가 높은 뉴스를 선택하여 읽을 수 있다.

[0029] 본 발명의 일실시예에 따르면, 현실 세계의 물리 객체의 외모와 같은 외적 특징이 아닌, 물리 객체의 의사 결정 특성과 같은 내적 특징이 반영된 디지털 트윈이 구현될 수 있다. 디지털 트윈의 외형은 다양한 형태로 표현될 수 있다.

[0030] 본 발명의 일실시예에 따른 강화 학습 기반의 가상 모델 학습 방법 및 가상 모델 서비스 제공 방법은, 프로세서 및 메모리를 포함하는 컴퓨팅 장치에서 수행될 수 있다.

[0031] 이하에서, 본 발명에 따른 실시예들을 첨부된 도면을 참조하여 상세하게 설명한다.

[0033] 도 1은 본 발명의 일실시예에 따른 강화 학습 기반의 가상 모델 학습 방법을 설명하기 위한 도면이며, 도 2는 본 발명의 일실시예에 따른 강화 학습 모델을 설명하기 위한 도면이다.

[0034] 도 1을 참조하면, 본 발명의 일실시예에 따른 컴퓨팅 장치는 선택지 집합에서, 타겟 롤모델에게 제공할 적어도 하나의 타겟 선택지를 결정(S110)한다. 여기서 선택지 집합은, 타겟 롤모델의 의사 결정의 대상이 될 수 있는 선택지들의 집합으로서, 예컨대, 다양한 종류의 뉴스들일 수 있다.

[0035] 그리고 컴퓨팅 장치는 타겟 선택지에 대한, 타겟 롤모델의 선호도 피드백 정보를 제공받는다(S120). 선호도 피드백 정보는 타겟 선택지에 대한 타겟 롤모델의 선택 유무에 따라 결정되는 정보로서, 타겟 선택지 중에서, 타겟 롤모델에 의해 선택된 선택지에 대해서는 제1보상값이 할당되고, 타겟 롤모델에 의해 선택되지 않은 선택지에 대해서는 제2보상값이 할당될 수 있다. 예컨대, 제1보상값으로 1이 할당되고, 제2보상값으로 0이 할당될 수 있다.

[0036] 컴퓨팅 장치는 선호도 피드백 정보을 이용하여, 타겟 롤모델에 대한 가상 모델의 학습을 수행(S130)하며, 가상 모델은, 입력 데이터에 대한 선호도값을 출력하는 강화 학습 기반의 선호도 네트워크를 포함한다.

[0037] 컴퓨팅 장치는 강화 학습 모델을 이용하여 가상 모델에 대한 학습을 수행할 수 있다. 본 발명의 일실시예에 따른 강화 학습 모델은 DQN(Deep Q-Network) 알고리즘 기반의 모델일 수 있으며, 도 2에 도시된 바와 같이, 에이전트(agent, 210)와 환경(environment, 220)으로 구성된다. 에이전트(210)는 선호도 네트워크(preference network, 211) 및 리플레이 메모리(replay memory, 212)를 포함하며, 환경(220)은 타겟 롤모델(221)을 포함한다.

[0038] 선호도 네트워크(211)는 DQN 알고리즘의 Q네트워크에 대응되며, 선호도값은 Q밸류에 대응된다. 그리고 리플레이 메모리(212)는 학습에 필요한 데이터를 저장한다.

[0039] 에이전트(210)는 환경(220)으로부터 제공된 상태 정보와 보상(reward) 정보인 선호도 피드백 정보를 기반으로 선호도 네트워크(211)에 대한 학습을 수행하며, 타겟 선택지라는 액션 정보를 출력한다. 여기서 상태 정보는, 타겟 롤모델(221)의 의사 결정과 관련된 정보로서, 타겟 롤모델(221)에 대한 특징 정보, 타겟 롤모델(221)에 의해 수행된 의사 결정에 대한 특징 정보 및 타겟 롤모델(221)의 의사 결정 시점에 대한 컨텍스트 정보를 포함할 수 있다.

[0040] 컴퓨팅 장치는 단계 S110 내지 S130을 반복하며, 가상 모델에 대한 학습을 수행하며, 이하 가상 모델 학습 방법의 각 단계별로 자세히 설명하기로 한다.

[0042] **타겟 선택지 결정(S110)**

[0043] 본 발명의 일실시예에 따른 컴퓨팅 장치는 타겟 롤모델(221)에 대한 특징 정보, 타겟 롤모델(221)에 의해 수행된 의사 결정에 대한 특징 정보 및 타겟 롤모델(221)의 의사 결정 시점에 대한 컨텍스트 정보를 이용하여, 선택지 집합에서 타겟 선택지를 결정한다.

[0044] 타겟 롤모델(221)에 대한 특징 정보는 타겟 롤모델(221)의 성별, 나이 및 국적 중 적어도 하나를 포함할 수 있다. 그리고, 타겟 롤모델(221)에 의해 수행된 의사 결정에 대한 특징 정보는, 미리 설정된 카테고리의 선택지 중에서, 타겟 롤모델(221)의 의사 결정에 의해 선택된 선택지에 대한 특징 정보를 포함한다. 일예로 미리 설정된 카테고리의 선택지가 뉴스라면, 의사 결정에 대한 특징 정보는, 타겟 롤모델(221)의 의사 결정에 의해 타겟 롤모델(221)이 읽고 싶어하는 것으로 선택된 뉴스의 카테고리, 보도 시간, 키워드 및 언론사 중 적어도 하나를 포함할 수 있다. 마지막으로 컨텍스트 정보는, 의사 결정 시점의 날짜, 시간 및 날씨 중 적어도 하나를 포함할

수 있다.

[0045] 여기서, 타겟 선택지의 결정 시점(time slot)을 t라고 할 경우, 타겟 롤모델(221)에 대한 특징 정보와, 타겟 롤모델(221)의 의사 결정에 의해 선택된 선택지에 대한 특징 정보는 h_t, 컨텍스트 정보는 c_t, 선택지 집합은 D_t로 정의될 수 있다. 이러한 정보와 선택지 집합은, 저장 장치에 저장된 상태에서 컴퓨팅 장치에 의해 로딩될 수 있다.

[0046] 선택지 집합에 포함된 선택지 중 하나를 d_t^n 으로 정의할 경우, d_t^n 은 [수학식 1]과 같이 표현될 수 있다.

수학식 1

[0047] $$d_t^n \in D_t = \left\{d_t^1, d_t^2, \ldots d_t^{N_l}\right\}$$

[0048] 여기서, N_l 은 선택지 집합에 포함된 선택지의 개수를 나타낸다.

[0049] 컴퓨팅 장치는 선택지 집합에 포함된 선택지, 타겟 롤모델(221)에 대한 특징 정보, 타겟 롤모델(221)에 의해 수행된 의사 결정에 대한 특징 정보 및 컨텍스트 정보를 선호도 네트워크(211)에 입력하여, 선택지 집합에 포함된 선택지 각각에 대한 선호도값을 획득한다. 그리고 선호도값에 따라서 타겟 선택지를 결정한다.

[0050] 타겟 선택지에 대한 선호도값을 y_t^n 으로 정의할 경우, 선호도 네트워크(211)가 출력하는 선호도값은, [수학식 2]로 표현될 수 있다.

수학식 2

[0051] $$y_t^n \in y_t = \left\{y_t^1, y_t^1, \ldots y_t^{N_l}\right\}$$

[0052] 컴퓨팅 장치는 일실시예로서, 학습 편향을 감소시키고 학습의 탐험(exploration)을 보장하기 위해, ε-탐욕(ε-greedy) 정책에 따라, 타겟 선택지를 결정할 수 있다. 즉, 컴퓨팅 장치는 입실론(ε) 확률로 선택지 집합에서 랜덤하게 타겟 선택지를 결정하고, 1-ε 확률로 선택지 집합에서 선도도값이 높은 순서로 타겟 선택지를 결정할 수 있다. 이를 통해 컴퓨팅 장치는 미리 설정된 개수(N_f)만큼의 타겟 선택지를 결정할 수 있다.

[0053] 타겟 선택지(l_t^k)의 리스트를 L_t로 정의할 경우, 타겟 선택지는 [수학식 3]과 같이 표현될 수 있다.

수학식 3

[0054] $$l_t^k \in L_t = \left\{l_t^1, l_t^2, \ldots l_t^{N_f}\right\}$$

[0055] 컴퓨팅 장치는 현재 타겟 선택지의 결정 시점(t) 이후인 다음 결정 시점(t+1)에서는, 새로운 선택지 집합(D_{t+1})을 대상으로, 새로운 타겟 롤모델에 대한 특징 정보 및 의사 결정에 대한 특징 정보(h_{t+1}), 컨

텍스트 정보(c_{t+1})를 이용해, 타겟 선택지를 결정한다.

[0057] **선호도 피드백 정보(S120)**

[0058] 타겟 롤모델(221)은, 타겟 선택지에 대해 자신의 선호도에 기반하여 의사 결정을 진행한다. 여기서, 의사 결정이란, 타겟 선택지의 선택 또는 미선택에 대한 의사 결정을 나타낸다. 전술된 바와 같이, 타겟 롤모델(221)이 선택한 타겟 선택지에 대해서는 제1보상값이 할당되고, 타겟 롤모델(221)이 선택하지 않은 타겟 선택지에 대해서는 제2보상값이 할당된다. 그리고 타겟 선택지에 대한 제1 및 제2보상값이 선호도 피드백 정보(r_t^k)로서, 에이전트(210)로 제공된다.

[0059] 컴퓨팅 장치는, 선호도 피드백 정보(r_t^k)와 함께, 타겟 롤모델(221)에 대한 특징 정보 및 의사 결정에 대한 특징 정보(h_t) 그리고 컨텍스트 정보(c_t)를 리플레이 메모리(212)에 저장한다. 그리고 후술되는 선호도 네트워크(211)에 대한 학습을 위해, 현재 타겟 선택지의 결정 시점(t) 이후인 다음 결정 시점(t+1)에 대한 선택지 집합(D_{t+1}), 타겟 롤모델(221)에 대한 특징 정보 및 의사 결정에 대한 특징 정보(h_{t+1}), 컨텍스트 정보(c_{t+1})를 리플레이 메모리(212)에 함께 저장한다.

[0061] **선호도 네트워크 학습(S130)**

[0062] 컴퓨팅 장치는 리플레이 메모리(212)에 저장된 정보($h_t,\ c_t,\ l_t^k,\ h_{t+1},\ c_{t+1},\ D_{t+1}\ ,r_t^k$)를 이용해, 선호도 네트워크(211)에 대한 학습을 수행한다. 이 때, 컴퓨팅 장치는 리플레이 메모리(212)에 저장된 정보를 미니 배치(mini-batch)로 나누어 학습을 수행할 수 있다.

[0063] 컴퓨팅 장치는, 타겟 선택지, 타겟 롤모델(221)에 대한 특징 정보, 의사 결정에 대한 특징 정보 및 컨텍스트 정보를 선호도 네트워크(211)에 입력하여, 타겟 선택지에 대한 선호도값을 획득하고, 선호도 피드백 정보 및 선호도값으로부터 계산된 손실값이 최소가 되도록, 선호도 네트워크(211)에 대한 학습을 수행한다.

[0064] 컴퓨팅 장치는 [수학식 4]와 같은 손실 함수를 이용해, 손실값을 계산하고, 이러한 손실값이 최소가 되도록 선호도 네트워크(211)의 가중치(Θ)를 업데이트할 수 있다.

수학식 4

[0065]
$$\|y^{TARGET} - P(h,c,l;\theta)\|^2$$

[0066] 여기서, y^{TARGET} 은 DQN 알고리즘의 타겟값(target value)을 나타낸다. $P(h,c,l;\theta)$ 는 선호도 네트워크(211)에서 출력되는 선호도값으로, y_t 에 대응되며, 타겟 선택지(l_t^k)에 대한 선호도값에 대응된다.

[0067] DQN 알고리즘의 학습 과정에서는 Q 네트워크이외 타겟 네트워크(target network)로부터 출력된 타겟값(target value)이 손실값 계산에 이용되며, 타겟값은 [수학식 5]와 같이 계산될 수 있다.

수학식 5

$$y^{TARGET} = r + \gamma P(h', c', \arg\max_{d \in D} P(h', c', d'; \theta); \theta')$$

[0068]

[0069] 여기서, h', c'는 전술된 h_{t+1}, c_{t+1}에 대응되며, d'는 D_{t+1}로부터 결정된 타겟 선택지에 대응된다. $P(h', c', d'; \theta)$는 d'에 대한 선호도값을 나타내며, θ'은 타겟 네트워크의 가중치를 나타낸다. γ는 디스카운트 팩터(discount factor)를 나타내며, r은 전술된 r_t^k에 대응된다. 타겟 네트워크의 가중치는 주기적으로 선호도 네트워크(211)의 가중치로 변경되면서 업데이트된다.

[0070] 본 발명의 일실시예에 따른 컴퓨팅 장치는 전술된 S110 내지 S130을 반복적으로 수행하며, 선호도 네트워크(211)에 대한 학습을 수행하며, 이러한 학습 방법에 대한 의사코드는 도 3과 같다.

[0072] 도 4는 본 발명의 일실시예에 따른 강화 학습 기반의 가상 모델 서비스 제공 방법을 설명하기 위한 도면이며, 도 5는 본 발명의 일실시예에 따른 가상 모델이 챗봇 형태로 서비스되는 일실시예를 나타내는 도면이다.

[0073] 도 4를 참조하면 본 발명의 일실시예에 따른 컴퓨팅 장치는 사용자의 요청에 따라, 저장 장치에 저장된 타겟 롤모델에 대한 특징 정보, 타겟 롤모델에 의해 수행된 의사 결정에 대한 특징 정보 및 사용자의 요청 시점에 대한 컨텍스트 정보를 로딩(S410)한다. 사용자의 요청 시점에 대한 컨텍스트 정보는, 사용자의 선호도값 요청 시점에서의 날짜, 시간 및 날씨 중 적어도 하나를 포함할 수 있다.

[0074] 그리고 컴퓨팅 장치는 타겟 선택지 및 로딩된 정보를, 타겟 롤모델에 대한 가상 모델에 입력하여, 타겟 선택지에 대한 가상 모델의 선호도값을 생성(S420)한다. 전술된 바와 같이, 가상 모델은 타겟 선택지에 대한 선호도값을 제공하는, 강화 학습 기반의 선호도 네트워크를 포함한다. 선호도 네트워크는 타겟 선택지 및 로딩된 정보를 입력받는 입력층, 은닉층 및 타겟 선택지에 대한 선호도값을 출력하는 출력층을 포함하는 인공 신경망일 수 있다.

[0075] 선호도값은 실시예에 따라서, 소프트맥스(softmax) 함수를 통해, 타겟 롤모델이 타겟 선택지를 선택할 확률 형태로 사용자에게 제공될 수 있으며, 사용자는 선호도값을 의사 결정에 활용할 수 있다.

[0076] 컴퓨팅 장치는 단계 S420에서 생성된 선호도값을 타겟 선택지와 매칭하여, 사용자에게 제공하거나, 또는 타겟 선택지 중에서, 선호도값에 따라 선택된 선택지를 사용자에게 제공할 수 있다. 컴퓨팅 장치는 선호도값이 높은 순서로 미리 설정된 개수만큼의 타겟 선택지를 사용자에게 제공할 수 있다.

[0077] 한편, 도 5에 도시된 바와 같이, 본 발명의 일실시예에 따른 가상 모델이 챗봇 형태로 구현될 수 있으며, 도 5에는 사용자가 타겟 선택지인 뉴스들에 대한 타겟 롤모델의 선호도를 뉴리라는 챗봇으로부터 제공받는 실시예가 도시된다.

[0078] 사용자가 챗봇에게 뉴스들에 대한 타겟 롤모델의 선호도값을 요청하면, 컴퓨팅 장치는 뉴스를 실시간으로 크롤링하고, 뉴스에 대한 키워드 등의 특징 정보를 추출할 수 있다. 그리고 뉴스에 대한 특징 정보가 타겟 선택지로서, 선호도 네트워크로 입력될 수 있다. 컴퓨팅 장치는 뉴스에 대한 선호도값 중에서 선호도값이 높은 순서로 미리 설정된 개수만큼의 뉴스를 선택하여, 사용자에게 제공할 수 있다.

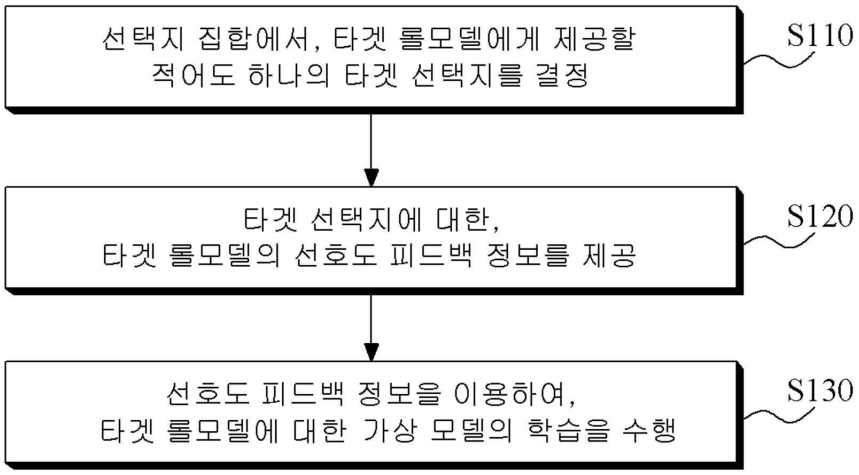
[0080] 앞서 설명한 기술적 내용들은 다양한 컴퓨터 수단을 통하여 수행될 수 있는 프로그램 명령 형태로 구현되어 컴퓨터 판독 가능 매체에 기록될 수 있다. 상기 컴퓨터 판독 가능 매체는 프로그램 명령, 데이터 파일, 데이터 구조 등을 단독으로 또는 조합하여 포함할 수 있다. 상기 매체에 기록되는 프로그램 명령은 실시예들을 위하여 특별히 설계되고 구성된 것들이거나 컴퓨터 소프트웨어 당업자에게 공지되어 사용 가능한 것일 수도 있다. 컴퓨터 판독 가능 기록 매체의 예에는 하드 디스크, 플로피 디스크 및 자기 테이프와 같은 자기 매체(magnetic

media), CD-ROM, DVD와 같은 광기록 매체(optical media), 플롭티컬 디스크(floptical disk)와 같은 자기-광 매체(magneto-optical media), 및 롬(ROM), 램(RAM), 플래시 메모리 등과 같은 프로그램 명령을 저장하고 수행하도록 특별히 구성된 하드웨어 장치가 포함된다. 프로그램 명령의 예에는 컴파일러에 의해 만들어지는 것과 같은 기계어 코드뿐만 아니라 인터프리터 등을 사용해서 컴퓨터에 의해서 실행될 수 있는 고급 언어 코드를 포함한다. 하드웨어 장치는 실시예들의 동작을 수행하기 위해 하나 이상의 소프트웨어 모듈로서 작동하도록 구성될 수 있으며, 그 역도 마찬가지이다.

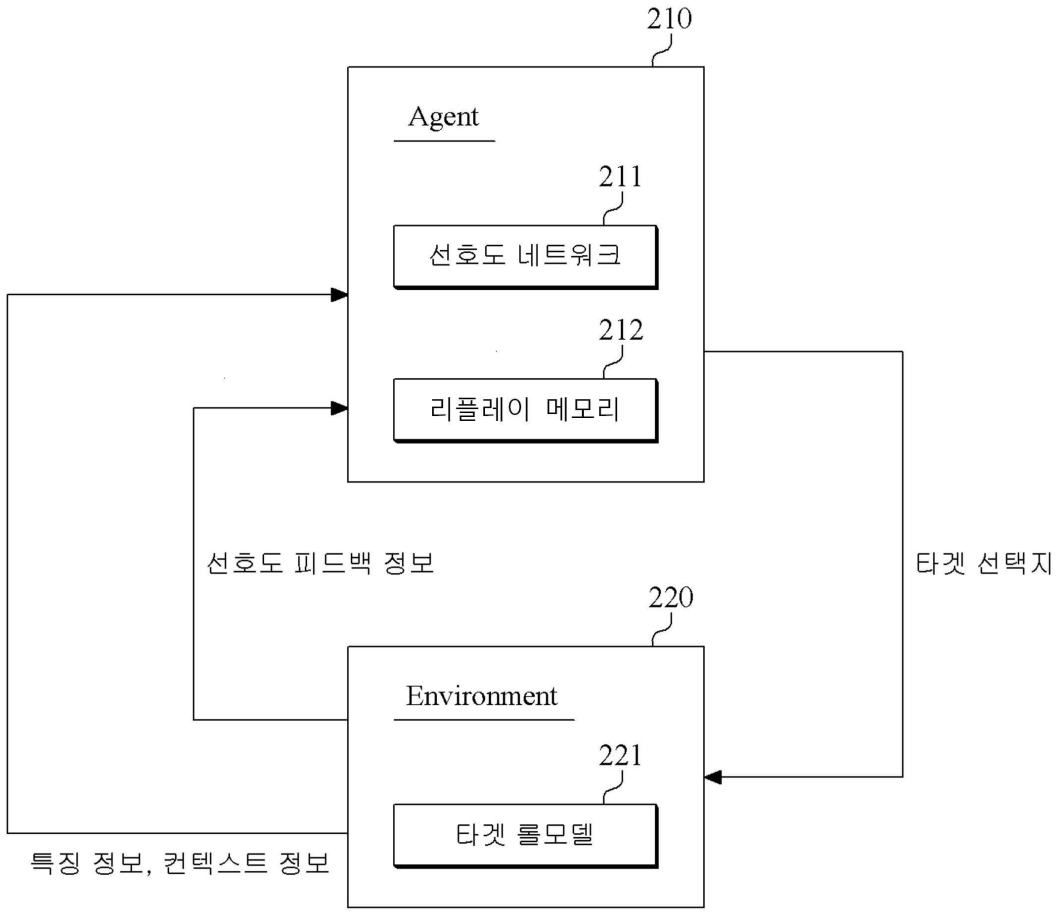
[0082] 이상과 같이 본 발명에서는 구체적인 구성 요소 등과 같은 특정 사항들과 한정된 실시예 및 도면에 의해 설명되었으나 이는 본 발명의 보다 전반적인 이해를 돕기 위해서 제공된 것일 뿐, 본 발명은 상기의 실시예에 한정되는 것은 아니며, 본 발명이 속하는 분야에서 통상적인 지식을 가진 자라면 이러한 기재로부터 다양한 수정 및 변형이 가능하다. 따라서, 본 발명의 사상은 설명된 실시예에 국한되어 정해져서는 아니되며, 후술하는 특허청구범위뿐 아니라 이 특허청구범위와 균등하거나 등가적 변형이 있는 모든 것들은 본 발명 사상의 범주에 속한다고 할 것이다.

도면

도면1



도면2


210
Agent
211
선호도 네트워크
212
리플레이 메모리
선호도 피드백 정보
타겟 선택지
220
Environment
221
타겟 롤모델
특징 정보, 컨텍스트 정보

도면3

Algorithm 1 Procedure of the Digital Role-model Twin generation

Notations : discount factor γ, epsilon ϵ, replay memory B, replay memory maximum size N_r, network parameter θ, target network parameter θ', role-model feature h, context feature c, decision feature D, decision reward r, preference list L, training batch size N_b, target network replacement frequency N'

Initialize θ

Initialize $\theta' \leftarrow \theta$

Initialize empty B

for time-slot $t \in \{1, 2, ...\}$ **do**

  Obtain h_t, c_t, $D_t \in D$

  for $n \in \{1, 2...N_l\}$ **do** ▷ in parallel

    Set $d_t^n \in D_t = \{d_t^1, d_t^2, ...d_t^{N_l}\}$

    Calculate output $y_t^n \in \{y_t^1, y_t^2, ...y_t^{N_l}\}$ using h_t, c_t, d_t^n

    $y_t^n = P(h_t, c_t, d_t^n)$

  end for

  for $k \in \{1, 2, ...N_f\}$ **do**

$$d^k = \begin{cases} \text{Randomly choose } d^k \text{ from } D_t & \text{with probability } \epsilon \\ \text{Choose } d^k \text{ with the largest } y_t^k \text{ from } D_t & \text{otherwise} \end{cases}$$

    Remove d^k from D_t

    $L_t \leftarrow L_t \cup \{d^k\}$

  end for

  Obtain c_{t+1} and $D_{t+1} \in D$ in next time-slot $t+1$

  for $k \in \{1, 2, ...N_f\}$ **do**

    Send $l_t^k \in L_t = \{l_t^1, l_t^2, ...l_t^{N_f}\}$ to Role-model

    Receive feedback of l_t^k from Role-model $r_t^k \in \{0, 1\}$

  end for

  Observe h_{t+1} in next time-slot $t+1$ from the feedback r_t^k of l_t^k's

  for $k \in \{1, 2, ...N_f\}$ **do**

    Store experience transition tuples $(h_t, c_t, l_t^k, h_{t+1}, c_{t+1}, D_{t+1}, r_t^k)$ to B, replacing the oldest tuple if $|B| \geq N_r$

  end for

  Sample a mini-batch of N_b tuples $(h_j, c_j, l_j, h_{j+1}, c_{j+1}, D_{j+1}, r_j) \sim \text{Unif}(B)$

  Construct target values, one for each of the N_b tuples:

  Define d' that gives maximum future reward is selected according to parameter θ

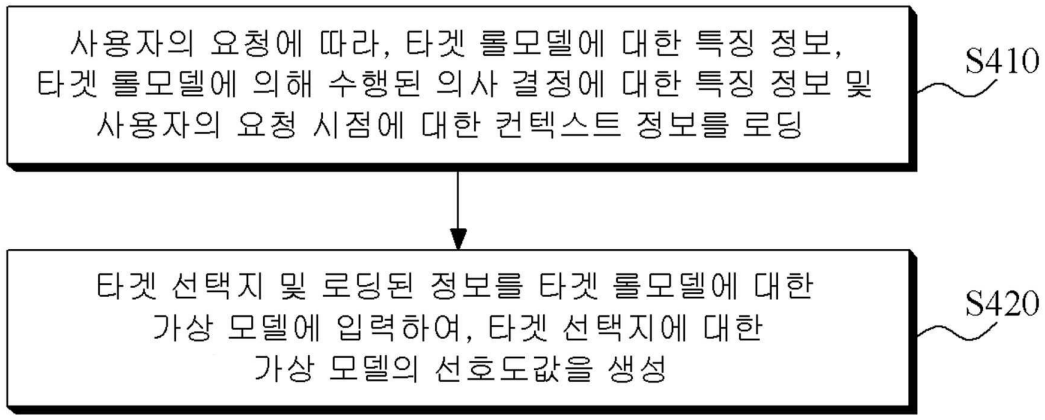
  $y_j = r_j + \gamma P(h_{j+1}, c_{j+1}, \underset{d' \in D_{j+1}}{\text{argmax}}\, P(h_{j+1}, c_{j+1}, d'; \theta); \theta')$

  Do a gradient descent step with loss $||y_j - P(h_j, c_j, l_j; \theta)||^2$ w.r.t. the network parameter θ

  Replace target parameters θ' as θ every N' step

end for

도면4



도면5

【심사관 직권보정사항】

【직권보정 1】

【보정항목】 청구범위

【보정세부항목】 청구항 10

【변경전】

컴퓨팅 장치에 의해 수행되는, 강화 학습 기반의 가상 모델 서비스 제공 방법에 있어서,

사용자의 요청에 따라, 저장 장치에 저장된 타겟 롤모델에 대한 특징 정보, 상기 타겟 롤모델에 의해 수행된 의사 결정에 대한 특징 정보 및 상기 사용자의 선호도값 요청 시점에 대한 컨텍스트 정보를 로딩하는 단계; 및

타겟 선택지 및 상기 로딩된 정보를 상기 타겟 롤모델에 대한 가상 모델에 입력하여, 상기 타겟 선택지에 대한 상기 가상 모델의 선호도값을 생성하는 단계를 포함하며,

상기 가상 모델은, 상기 타겟 선택지에 대한 선호도값을 제공하는, 강화 학습 기반의 선호도 네트워크를 포함하며,

상기 컨텍스트 정보는

상기 선호도값 의사 결정 시점의 날짜, 시간 및 날씨 중 적어도 하나를 포함하는

강화 학습 기반의 가상 모델 서비스 제공 방법.

【변경후】

컴퓨팅 장치에 의해 수행되는, 강화 학습 기반의 가상 모델 서비스 제공 방법에 있어서,

사용자의 요청에 따라, 저장 장치에 저장된 타겟 롤모델에 대한 특징 정보, 상기 타겟 롤모델에 의해 수행된 의사 결정에 대한 특징 정보 및 상기 사용자의 선호도값 요청 시점에 대한 컨텍스트 정보를 로딩하는 단계; 및

타겟 선택지 및 상기 로딩된 정보를 상기 타겟 롤모델에 대한 가상 모델에 입력하여, 상기 타겟 선택지에 대한 상기 가상 모델의 선호도값을 생성하는 단계를 포함하며,

상기 가상 모델은, 상기 타겟 선택지에 대한 선호도값을 제공하는, 강화 학습 기반의 선호도 네트워크를 포함하며,

상기 컨텍스트 정보는

선호도값 의사 결정 시점의 날짜, 시간 및 날씨 중 적어도 하나를 포함하는

강화 학습 기반의 가상 모델 서비스 제공 방법.