(19) 대한민국특허청(KR)
(12) 등록특허공보(B1)

(45) 공고일자 2023년05월30일
(11) 등록번호 10-2538455
(24) 등록일자 2023년05월25일

(51) 국제특허분류(Int. Cl.)
H04N 21/466 (2011.01) *G06N 20/00* (2019.01)
G06Q 50/10 (2012.01) *H04N 21/44* (2011.01)
H04N 21/45 (2011.01)
(52) CPC특허분류
H04N 21/4668 (2013.01)
G06N 20/00 (2021.08)
(21) 출원번호 10-2022-0114848
(22) 출원일자 2022년09월13일
심사청구일자 2022년09월13일
(56) 선행기술조사문헌
KR1020180042934 A*
KR1020210102727 A*
KR102415719 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)
이다은
서울특별시 광진구 군자로10길 17, 105동 202호(군자동, 대명이튼캐슬)
(74) 대리인
민영준

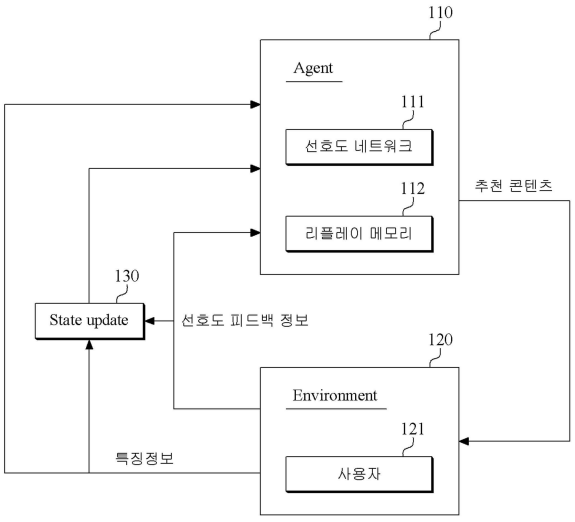
전체 청구항 수 : 총 9 항 심사관 : 장우진

(54) 발명의 명칭 **강화 학습 기반의 콘텐츠 추천을 위한 학습 방법 및 콘텐츠 추천 방법**

(57) 요 약

강화 학습 기반의 콘텐츠 추천을 위한 학습 방법 및 콘텐츠 추천 방법이 개시된다. 개시된, 강화 학습 기반의 콘텐츠 추천을 위한 학습 방법은 사용자의 컨텐츠 선호도에 대한 제1특징 정보, 추천 콘텐츠에 대한 제2특징 정보 및 상기 추천 콘텐츠에 대한 사용자의 선호도 피드백 정보를 이용하여, 상기 제1 및 제2특징 정보를 갱신하는 단계; 및 상기 제1 및 제2특징 정보 및 상기 갱신된 제1 및 제2특징 정보를 이용하여, 상기 추천 콘텐츠에 대한 사용자의 선호도값을 예측하는 선호도 네트워크에 대한 강화 학습을 수행하는 단계를 포함한다.

대 표 도 - 도1



(52) CPC특허분류
G06Q 50/10 (2015.01)
H04N 21/44008 (2013.01)
H04N 21/4532 (2013.01)
H04N 21/4662 (2013.01)

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

사용자의 컨텐츠 선호도에 대한 제1특징 정보, 추천 콘텐츠에 대한 제2특징 정보 및 상기 추천 콘텐츠에 대한 사용자의 선호도 피드백 정보를 이용하여, 상기 제1 및 제2특징 정보를 갱신하는 단계; 및

상기 제1 및 제2특징 정보 및 상기 갱신된 제1 및 제2특징 정보를 이용하여, 상기 추천 콘텐츠에 대한 사용자의 선호도값을 예측하는 선호도 네트워크에 대한 강화 학습을 수행하는 단계를 포함하며,

상기 선호도 네트워크에 대한 강화 학습을 수행하는 단계는

상기 제1 및 제2특징 정보를 상기 선호도 네트워크에 입력하여, 상기 추천 콘텐츠에 대한 선호도값을 획득하는 단계;

상기 갱신된 제1 및 제2특징 정보를 타겟 네트워크에 입력하 여, 상기 추천 콘텐츠에 대한 타겟값을 획득하는 단계; 및

상기 선호도값 및 타겟값으로부터 계산된 손실값이 최소가 되도록, 상기 선호도 네트워크에 대한 학습을 수행하는 단계

를 포함하는 강화 학습 기반의 콘텐츠 추천을 위한 학습 방법.

청구항 2

제 1항에 있어서,

상기 제1 및 제2특징 정보를 갱신하는 단계는

상기 제1 및 제2특징 정보 사이의 내적값과, 상기 선호도 피드백 정보에 포함된 보상값 사이의 차이에 따라서, 상기 제1 및 제2특징 정보를 갱신하는

강화 학습 기반의 콘텐츠 추천을 위한 학습 방법.

청구항 3

제 2항에 있어서,

상기 제1 및 제2특징 정보를 갱신하는 단계는

하기 수학식1을 이용하여, 상기 제1특징 정보를 갱신하는

강화 학습 기반의 콘텐츠 추천을 위한 학습 방법.

[수학식 1]

$$s_{t+1} = s_t - 2\alpha[(s_t C_{a_t}^T - \mathrm{r})C_{a_t} + \gamma s_t]$$

여기서, s_t 는 상기 제1특징 정보, s_{t+1} 는 상기 갱신된 제1특징 정보, C_{a_t}는 상기 제2특징 정보, α는 학습률(learning rate), r은 상기 보상값, γ 는 미리 설정된 상수값을 나타냄.

청구항 4

제 2항에 있어서,

상기 제1 및 제2특징 정보를 갱신하는 단계는

하기 수학식2를 이용하여, 상기 제2특징 정보를 갱신하는

강화 학습 기반의 콘텐츠 추천을 위한 학습 방법.

[수학식 2]

$$C_{a_t} = C_{a_t} - 2\alpha[(C_{a_t} s_t^T - \mathrm{r})s_t + \gamma C_{a_t}]$$

s_t 는 상기 제1특징 정보, C_{a_t} 는 상기 제2특징 정보, α는 학습률(learning rate), r은 상기 보상값, γ 는 미리 설정된 상수값을 나타냄.

청구항 5

삭제

청구항 6

제 1항에 있어서,

상기 추천 콘텐츠는

후보 콘텐츠들 중에서, 상기 선호도값에 기반하여 결정된 콘텐츠인

강화 학습 기반의 콘텐츠 추천을 위한 학습 방법.

청구항 7

삭제

청구항 8

사용자의 컨텐츠 선호도에 대한 제1특징 정보 및 후보 콘텐츠들에 대한 제2특징 정보를 입력받는 단계;

상기 제1 및 제2특징 정보를 미리 강화 학습된 선호도 네트워크에 입력하여, 상기 후보 콘텐츠들에 대한 사용자의 선호도값을 획득하는 단계;

상기 선호도값에 기반하여, 상기 후보 콘텐츠들 중에서 추천 콘텐츠를 결정하는 단계;

상기 추천 콘텐츠에 대한 사용자의 선호도 피드백 정보를 이용하여, 상기 제1특징 정보 및 상기 추천 콘텐츠에 대한 제3특징 정보를 갱신하는 단계; 및

상기 제1 및 제3특징 정보 및 상기 갱신된 제1 및 제3특징 정보를 이용하여, 상기 선호도 네트워크에 대한 강화 학습을 수행하는 단계를 포함하며,

상기 선호도 네트워크에 대한 강화 학습을 수행하는 단계는

상기 제1 및 제3특징 정보를 상기 선호도 네트워크에 입력하여, 상기 추천 콘텐츠에 대한 선호도값을 획득하는 단계;

상기 갱신된 제1 및 제3특징 정보를 타겟 네트워크에 입력하여, 상기 추천 콘텐츠에 대한 타겟값을 획득하는 단계; 및

상기 선호도값 및 타겟값으로부터 계산된 손실값이 최소가 되도록, 상기 선호도 네트워크에 대한 학습을 수행하

는 단계

를 포함하는 강화 학습 기반의 콘텐츠 추천 방법.

청구항 9

삭제

청구항 10

제 8항에 있어서,

상기 제1 및 제3특징 정보를 갱신하는 단계는

상기 제1 및 제3특징 정보 사이의 내적값과, 상기 선호도 피드백 정보에 포함된 보상값 사이의 차이에 따라서, 상기 제1 및 제3특징 정보를 갱신하는

강화 학습 기반의 콘텐츠 추천 방법.

청구항 11

제 8항에 있어서,

상기 후보 콘텐츠들은

메타버스 환경에서 제공되는 콘텐츠들인

강화 학습 기반의 콘텐츠 추천 방법.

청구항 12

제 11항에 있어서,

상기 선호도 피드백 정보에 따라서, 상기 메타버스 환경을 변경하는 단계

를 더 포함하는 강화 학습 기반의 콘텐츠 추천 방법

발명의 설명

기 술 분 야

[0001] 본 발명은 강화 학습 기반의 콘텐츠 추천을 위한 학습 방법 및 콘텐츠 추천 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 융복합된 디지털 세계, 초월 세계를 의미한다. 메타버스는 새로운 소셜 플랫폼으로 떠오르고 있으며, 가상과 현실을 아우르며 사회적, 경제적, 문화적 활동이 가능한 새로운 융합 소셜 플랫폼이다.

[0004] 기존 메타버스 플랫폼은 특정한 게임의 목적으로만 사용되었다면, 비대면 생활이 일상화가 되면서 교육, 경제, 문화 등 다양한 분야에서 메타버스 플랫폼을 활용하여 소비자와의 소통할 수 있는 플랫폼 구축하고 있다. 메타버스 전시회, 메타버스 회의, 메타버스 콘서트 등 메타버스 플랫폼 자체에 대한 활용 방법은 다양해지고 있으나, 메타버스 사용자의 경험 만족도를 높이고자 하는 연구는 많이 이뤄지고 있지 않다.

[0005] 플랫폼에 대한 사용자의 경험 만족도는 만족도 조사와 추천 의향으로 확인할 수 있다. 만족도는 사용자의 전반적인 만족도를 확인하는 지표이며, 추천 의향은 만족감을 통해 다른 사용자에게 플랫폼을 추천하고자 하는 것을

의미한다.

[0006] 메타버스 콘텐츠 경험에서 만족도와 추천 의향에 영향을 미치는 요인은 메타버스 어포던스 디자인 요소를 바탕으로 분석할 수 있다. 어포던스 디자인은 어떠한 하나의 행동을 유발하는 디자인이다. 예컨대, 가위의 손잡이 모양은, 올바른 가위 사용법을 유도하며, 이때의 가위 손잡이 모양이 어포던스 디자인 요소이다.

[0007] 어포던스 디자인은 디지털 사회가 발전함에 따라 디지털 환경 디자인 측면에서도 다수 적용되어 사용자 경험 만족도를 높이는 방법으로 활용되고 있다. 디지털 환경에서의 어포던스 디자인은 감각적 요소, 기능적 요소, 지각적 요소, 총 3가지로 분류한다.

[0008] 감각적 요소는 가상세계에서의 시간과 현실 세계에서의 시간상의 오차를 최대한 줄여 동일한 시간상의 감각을 느낄 수 있게 하는 요소이다. 기능적 요소는 가상세계에서의 행동과 현실 세계에서의 행동이 유사하게 될 수 있도록 하는 조작 요소이며, 마지막으로 지각적 요소는 가상세계에서의 콘텐츠와 사용자의 상호작용을 통해 지속적인 관계가 유지될 수 있도록 하는 요소이다.

[0009] 디지털 환경에서의 어포던스 디자인 요소에 대한 선행 연구에서는 어포던스 디자인의 3가지 요소 중 사용자의 경험 만족도에 가장 긍정적인 영향을 주는 요소는 지각적 요소라고 말한다. 콘텐츠와 사용자의 상호작용을 높이는 역할을 하는 지각적인 요소는 상호작용이 활발하게 일어날 수 있도록 하는 요소를 말한다. 이에 대한 예시로는 메타버스 플랫폼 중 하나인 '로블록스'를 들 수 있다.

[0010] '로블록스'는 사용자가 게임을 만들어 다른 사용자들도 즐길 수 있도록 하는 온라인 게임 플랫폼이다. 사용자가 설정하는 것에 따라 가상공간 내 요소들을 변화시킴으로써 메타버스 내 콘텐츠와 사용자 간 상호작용이 발생하게 된다. 앞선 '로블록스' 예시와 같이 비교적 많은 활용이 이뤄졌던 게임 분야에서는 다양한 요소들로 사용자의 경험 만족도를 높이고자 하는 연구들이 많이 진행되었다.

[0011] 하지만, 기존 연구들은 게임 분야에만 집중됐을 뿐만 아니라 메타버스만의 현실과 가상세계를 아우르는 특징을 고려하기보단 '가상공간'에 맞춘 연구들이 대부분이었다. 메타버스 콘텐츠의 사용자 경험 만족도를 개선하기 위해선 메타버스만의 특징을 활용한 연구가 필요하다.

[0012] 관련 선행문헌으로 대한민국 등록특허 제10-2329074호, 대한민국 공개특허 제2021-0157337호, 제2021-0074246호가 있다.

발명의 내용

해결하려는 과제

[0014] 본 발명은 강화 학습 모델을 이용하여, 사용자가 선호하는 콘텐츠를 추천하는 방법을 제공하기 위한 것이다.

[0015] 특히 본 발명은 메타버스 환경에서, 사용자가 선호하는 콘텐츠를 추천하는 방법을 제공하기 위한 것이다.

과제의 해결 수단

[0017] 상기한 목적을 달성하기 위한 본 발명의 일 실시예에 따르면, 사용자의 컨텐츠 선호도에 대한 제1특징 정보, 추천 콘텐츠에 대한 제2특징 정보 및 상기 추천 콘텐츠에 대한 사용자의 선호도 피드백 정보를 이용하여, 상기 제1 및 제2특징 정보를 갱신하는 단계; 및 상기 제1 및 제2특징 정보 및 상기 갱신된 제1 및 제2특징 정보를 이용하여, 상기 추천 콘텐츠에 대한 사용자의 선호도값을 예측하는 선호도 네트워크에 대한 강화 학습을 수행하는 단계를 포함하는 강화 학습 기반의 콘텐츠 추천을 위한 학습 방법이 제공된다.

[0018] 또한 상기한 목적을 달성하기 위한 본 발명의 다른 실시예에 따르면, 사용자의 컨텐츠 선호도에 대한 제1특징 정보 및 후보 콘텐츠들에 대한 제2특징 정보를 입력받는 단계; 상기 제1 및 제2특징 정보를 미리 강화 학습된 선호도 네트워크에 입력하여, 상기 후보 콘텐츠들에 대한 사용자의 선호도값을 획득하는 단계; 및 상기 선호도값에 기반하여, 상기 후보 콘텐츠들 중에서 추천 콘텐츠를 결정하는 단계를 포함하는 강화 학습 기반의 콘텐츠 추천 방법이 제공된다.

발명의 효과

[0020] 본 발명의 일실시예에 따르면, 콘텐츠에 대한 사용자의 피드백 정보에 의해 사용자의 콘텐츠 선호도 정보가 실시간으로 갱신됨으로써, 사용자가 보다 선호할 수 있는 콘텐츠가 사용자에게 추천될 수 있다.

[0021] 또한 본 발명의 일실시예에 따르면, 메타버스 콘텐츠와 사용자의 상호 작용을 통해 지각적 어포던스 디자인 요소가 개선될 수 있으며, 이를 통해 사용자의 경험 만족도가 더욱 개선될 수 있다.

도면의 간단한 설명

[0023] 도 1은 본 발명의 일실시예에 따른 강화 학습 모델을 설명하기 위한 도면이다.

도 2는 본 발명의 일실시예에 따른 강화 학습 기반의 콘텐츠 추천을 위한 학습 방법을 설명하기 위한 도면이다.

도 3은 본 발명의 일실시예에 따른 특징 정보 갱신 방법의 의사 코드를 나타내는 도면이다.

도 4는 본 발명의 일실시예에 따른 선호도 네트워크 학습 방법을 설명하기 위한 도면이다.

도 5는 본 발명의 일실시예에 따른 선호도 네트워크 학습 방법의 의사 코드를 나타내는 도면이다.

도 6은 본 발명의 일실시예에 따른 강화 학습 기반의 콘텐츠 추천 방법을 설명하기 위한 도면이다.

도 7은 본 발명의 구체적 실시예에 따른 강화 학습 기반의 콘텐츠 추천 방법을 설명하기 위한 도면이다.

발명을 실시하기 위한 구체적인 내용

[0024] 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.

[0026] 콘텐츠 추천 시스템은, 게임 이외 메타버스만의 특징을 매우 잘 활용할 수 있는 어플리케이션이라고 할 수 있다. 메타버스의 시간적, 공간적 제약이 없는 환경에서는 콘텐츠들에 대한 접근이 용이한 반면, 무수히 많은 콘텐츠에 의해 사용자의 경험 만족도는 낮아질 수 있다.

[0027] 따라서, 사용자가 선호할만한 콘텐츠를 추천하는 것은, 메타버스 환경에서 사용자의 경험 만족도를 높일 수 있는 방법이며, 본 발명은 강화 학습 기반으로 사용자에게 콘텐츠를 추천하는 방법을 제안한다.

[0028] 콘텐츠 추천을 위해, 협업 필터링(collaborative filtering, CF) 방법, 강화학습 방법 등이 연구되고 있다. 협업 필터링 방법의 경우, 콘텐츠 추천을 위해, 콘텐츠 추천 대상인 사용자 뿐만 아니라 또다른 사용자의 정보가 필수적으로 필요한 반면, 강화학습 방법의 경우, 콘텐츠 추천 대상인 사용자에 대한 정보만으로 콘텐츠 추천이 가능한 장점이 있다.

[0029] 본 발명의 일실시예에 따른 콘텐츠 추천 방법은, 메타버스 환경 뿐만 아니라, 콘텐츠의 추천이 필요한 다양한 환경에서 적용될 수 있으며, 프로세서 및 메모리를 포함하는 컴퓨팅 장치에서 수행될 수 있다.

[0030] 이하에서, 본 발명에 따른 실시예들을 첨부된 도면을 참조하여 상세하게 설명한다.

[0032] 도 1은 본 발명의 일실시예에 따른 강화 학습 모델을 설명하기 위한 도면이며, 도 2는 본 발명의 일실시예에 따른 강화 학습 기반의 콘텐츠 추천을 위한 학습 방법을 설명하기 위한 도면이다.

[0033] 본 발명의 일실시예에 따른 강화 학습 모델은 DQN(Deep Q-Network) 알고리즘 기반의 모델일 수 있으며, 도 1에 도시된 바와 같이, 에이전트(agent, 110)와 환경(environment, 120)으로 구성된다. 에이전트(110)는 선호도 네트워크(preference network, 111) 및 리플레이 메모리(replay memory, 112)를 포함하며, 환경(120)은 사용자(121)를 포함한다.

[0034] 선호도 네트워크(111)는 콘텐츠들에 대한 사용자의 선호도값을 예측하는 네트워크로서, 일실시예로서 인공 신경망일 수 있다. 선호도 네트워크(111)는 DQN 알고리즘의 Q네트워크에 대응되며, 선호도 네트워크(111)에서 출력되는 선호도값은 Q밸류에 대응된다. 그리고 리플레이 메모리(112)는 학습에 필요한 데이터를 저장한다.

[0035] 에이전트(110)는 환경(120)으로부터 제공된 상태(state) 정보와 보상(reward)값을 이용해, 선호도 네트워크(211)에 대한 학습을 수행하며, 후보 콘텐츠들 중에서 추천 콘텐츠를 결정하여, 추천 콘텐츠라는 액션(action) 정보를 출력한다. 여기서, 상태 정보는 사용자의 선호도에 대한 특징 정보와, 추천 콘텐츠에 대한 특징 정보를 포함하며, 이러한 특징 정보는 갱신(130)되어 에이전트(110)로 제공될 수 있다. 추천 콘텐츠에 대한 선호도 피드백 정보는 추천 콘텐츠를 경험한 사용자(121)에 의해 결정되며, 보상값을 포함한다. 그리고 콘텐츠에 대한 특징 정보는 저장 장치에 저장될 수 있다.

[0036] 본 발명의 일실시예에 따른 컴퓨팅 장치는 전술된 강화 학습 모델을 이용하여, 콘텐츠 추천을 위한 학습을 수행한다.

[0037] 도 2를 참조하면 본 발명의 일실시예에 따른 컴퓨팅 장치는 사용자(121)의 컨텐츠 선호도에 대한 제1특징 정보, 추천 콘텐츠에 대한 제2특징 정보 및 추천 콘텐츠에 대한 사용자(121)의 선호도 피드백 정보를 이용하여, 제1 및 제2특징 정보를 갱신(S120)한다. 여기서, 추천 콘텐츠는 후보 콘텐츠들 중에서, 선호도 네트워크(111)에서 출력되는 선호도값에 기반하여 결정된 콘텐츠로서, ε-탐욕(ε-greedy) 정책에 따라 결정될 수 있다. 그리고 선호도 피드백 정보는 사용자(121)가 추천 콘텐츠를 경험한 이후, 추천 콘텐츠에 대한 선호도를 포함하는 정보로서, 이러한 선호도는 전술된 보상값에 대응된다.

[0038] 그리고 컴퓨팅 장치는 제1 및 제2특징 정보 및 갱신된 제1 및 제2특징 정보를 이용하여, 추천 콘텐츠에 대한 사용자(121)의 선호도값을 예측하는 선호도 네트워크(111)에 대한 강화 학습을 수행(S220)한다.

[0039] 컴퓨팅 장치는 단계 S210 및 S220을 반복하며, 선호도 네트워크(111)에 대한 학습을 수행하며, 이하 학습 방법의 각 단계별로 자세히 설명하기로 한다.

[0040] 본 발명의 일실시예에 따르면, 콘텐츠에 대한 사용자의 피드백 정보에 의해 사용자의 콘텐츠 선호도 정보가 실시간으로 갱신됨으로써, 사용자가 보다 선호할 수 있는 콘텐츠가 사용자에게 추천될 수 있다.

[0042] **특징 정보 갱신(S210)**

[0043] 제1 및 제2특징 정보는 일실시예로서, 콘텐츠의 속성에 대한 특징값을 포함하는 미리 설정된 크기의 벡터일 수 있다. 일예로서, 콘텐츠가 미술 작품이며, 콘텐츠에 대한 속성이 적색 계열 속성, 녹색 계열 속성, 청색 계열 속성으로 정의될 수 있으며, 각 속성값은 0에서 1사이의 값으로 표현될 수 있다. 사용자가 적색 계열 속성을 선호하는 경우 제1특징 정보는 [1, 0, 0]과 같이 표현될 수 있다. 그리고 제2특징 정보는 미술 작품에 포함된 적색 계열 속성, 녹색 계열 속성, 청색 계열 속성의 비율에 따라서 결정될 수 있으며, 예컨대, 적색 계열 속성의 비율이 0.5, 녹색 계열 속성의 비율이 0.3, 청색 계열 속성의 비율이 0.2라면, 제2특징 정보는 [0.5, 0.3, 0.2]로 표현될 수 있다.

[0044] 선호도 피드백 정보에 포함된 보상값은 사용자가 입력하는 -1에서 1사이의 점수로 표현될 수 있으며, 보상값이 클수록 추천 콘텐츠에 대한 선호도가 높음을 의미한다. 또한 사용자가 입력하는 점수 뿐만 아니라, 사용자의 콘텐츠 경험 시간, 사용자가 콘텐츠를 경험할 때 나타나는 사용자의 표정 등이 보상값으로 이용될 수 있다.

[0045] 본 발명의 일실시예에 따른 컴퓨팅 장치는 제1 및 제2특징 정보 사이의 내적값과, 선호도 피드백 정보에 포함된 보상값 사이의 차이에 따라서, 제1 및 제2특징 정보를 갱신한다. 제1 및 제2특징 정보의 내적값이 크다는 것은, 제1 및 제2특징 정보 사이에 공통된 특징이 많음을 의미하며, 이는 사용자가 해당 콘텐츠에 대한 선호도가 크다는 것을 의미한다. 그리고 선호도 피드백 정보에 포함된 보상값은 사용자의 선호도에 비례하며, 따라서 내적값과 보상값의 차이가 크다는 것은 제1 및 제2특징 정보가, 사용자의 취향을 정확하게 반영하지 못하고 있음을 의미한다. 이에 컴퓨팅 장치는 내적값과 보상값의 차이가 감소하도록, 제1 및 제2특징 정보를 갱신할 수 있다.

[0046] 일예로서, 컴퓨팅 장치는 [수학식 1]을 이용하여 제1특징 정보(s_t)를 갱신할 수 있다.

수학식 1

[0047]
$$s_{t+1} = s_t - 2\alpha[(s_t C_{a_t}^T - r)C_{a_t} + \gamma s_t]$$

[0048] 여기서, s_t는 현재 시점(t)에서의 제1특징 정보를 나타내며, s_{t+1}는 다음 시점(t+1)에서의 갱신된 제1특징 정보를 나타낸다. 그리고 C_{a_t}는 제2특징 정보를 나타내며, α는 학습률(learning rate)를 나타낸다. 그리고 r은 보상값, γ는 미리 설정된 상수값을 나타낸다.

[0049] 또한 컴퓨팅 장치는 일예로서, [수학식 2]를 이용하여 제2특징 정보(C_{a_t})를 갱신할 수 있다.

수학식 2

[0050] $$C_{a_t} = C_{a_t} - 2\alpha[(C_{a_t} s_t^T - \mathrm{r})s_t + \gamma C_{a_t}]$$

[0051] 이 때, 컴퓨팅 장치는 과적합을 막기 위해, 갱신된 특징 정보에 대해 정규화를 진행할 수 있으며, [수학식 3]과 같은 목적함수를 만족하도록 특징 정보를 갱신할 수 있다.

수학식 3

[0052] $$\min_{s,\{C_i\}\forall i\in I} \sum_{i\in I} (s^\top C_i - r_{(s,i)})^2 + \gamma(\|s\|^2 + \|C_i\|^2)$$

[0053] 여기서, i는 추천 콘텐츠에 대한 인덱스를 나타낸다.

[0054] 이러한 특징 정보의 갱신 과정을 통해, 제1특징 정보는 실시간으로 사용자의 콘텐츠 선호도의 특징을 더욱 잘 반영하게 되며, 제2특징 정보는 사용자에 따른 특징을 나타내는 특징 정보로 갱신된다. 즉, 제2특징 정보는 사용자 별로 할당되는 추천 콘텐츠에 대한 특징 정보로서, 제2특징 정보는 사용자별로 갱신되는 정보로서, 사용자별로 개인화된 정보일 수 있다.

[0055] 전술된 특징 정보 갱신 과정을 의사 코드로 표현하면, 도 3과 같다.

[0056] 한편, 현재 시점에서의 제1 및 제2특징 정보와, 보상값, 갱신된 제1 및 제2특징 정보는 리플레이 메모리(112)에 저장된다. 그리고 후보 콘텐츠들 중 다른 콘텐츠들 역시 또다른 사용자에게 추천 콘텐츠로 제공되며, 이 과정에서 후보 콘텐츠들에 대한 제3특징 정보 역시 갱신될 수 있고, 갱신된 제3특징 정보 역시 리플레이 메모리(112)에 저장된다. 리플레이 메모리(112)에 저장된 정보들은, 선호도 네트워크(111)의 학습에 이용된다.

[0058] **선호도 네트워크 학습(S220)**

[0059] 컴퓨팅 장치는 리플레이 메모리(112)에 저장된 정보인, 제1특징 정보, 제2특징 정보, 보상값, 갱신된 제1특징 정보, 갱신된 제2 및 제3특징 정보를 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.

[0060] 도 4를 참조하면, 컴퓨팅 장치는 제1 및 제2특징 정보를 선호도 네트워크(111)에 입력하여, 추천 콘텐츠에 대한 선호도값을 획득(S221)한다. 선호도 네트워크(111)의 출력값은 [수학식 4]와 같이 표현될 수 있다.

수학식 4

[0061] $$Q(s_j, C_{a_j}^j; \theta)$$

[0062] 여기서, s_j 는 제1특징 정보, $C_{a_j}^j$ 는 제2특징 정보, θ 는 선호도 네트워크(111)의 가중치를 나타낸다. 즉, 선호도 네트워크는 제1 및 제2특징 정보를 입력받아, 추천 콘텐츠에 대한 선호도값을 출력한다.

[0063] 그리고 갱신된 제1 및 제2특징 정보를 타겟 네트워크에 입력하여, 추천 콘텐츠에 대한 타겟값을 획득(S222)한다. DQN 알고리즘의 학습 과정에서는 Q 네트워크이외 타겟 네트워크(target network)로부터 획득된 타겟값(target value)이 손실값 계산에 이용된다.

[0064] 컴퓨팅 장치는 일실시예로서, [수학식 5]를 이용하여 타겟값(y_j)을 계산할 수 있다.

수학식 5

[0065] $$y_j = r_j + \gamma \hat{Q}(s_{j+1}, C_{i^*}; \bar{\theta}))$$

[0066] 여기서, j는 미니 배치의 인덱스, $\bar{\theta}$ 는 타겟 네트워크의 가중치, γ 는 디스카운트 팩터(discount factor), s_{j+1} 은 갱신된 제1특징 정보를 나타낸다. 그리고 $\hat{Q}(s_{j+1}, C_{i^*}; \bar{\theta}))$ 는 타겟 네트워크가 출력하는 추천 및 후보 콘텐츠 중 최대 선호도값에 대응되는 콘텐츠(C_{i^*})의 선호도값을 나타낸다. 즉, 타겟 네트워크는 갱신된 제1특징 정보 및 갱신된 제2 및 제3특징 정보를 입력받아, 추천 및 후보 콘텐츠에 대한 선호도값을 출력한다.

[0067] 그리고 컴퓨팅 장치는 선호도값 및 타겟값으로부터 계산된 손실값이 최소가 되도록, 선호도 네트워크(111)에 대한 학습을 수행(S223)한다. 컴퓨팅 장치는 [수학식 6]과 같은 손실 함수를 이용해, 손실값을 계산할 수 있으며, 손실값이 최소가 되도록 선호도 네트워크(111)의 가중치(Θ)를 업데이트할 수 있다.

수학식 6

[0068] $$(y_j - Q(s_j, C^j_{a_j}; \theta))^2$$

[0069] 본 발명의 일실시예에 따른 컴퓨팅 장치는 전술된 S211 내지 S213을 반복적으로 수행하며, 선호도 네트워크(211)에 대한 학습을 수행하며, 이러한 학습 방법에 대한 의사코드는 도 5와 같다. 그리고 타겟 네트워크의 가중치는 주기적으로 선호도 네트워크(111)의 가중치로 변경되면서 업데이트된다. 또한 단계 S211 및 S212는 서로 병렬적으로 수행될 수 있다.

[0071] 도 6은 본 발명의 일실시예에 따른 강화 학습 기반의 콘텐츠 추천 방법을 설명하기 위한 도면이다.

[0072] 도 6을 참조하면, 본 발명의 일실시예에 따른 컴퓨팅 장치는 사용자의 컨텐츠 선호도에 대한 제1특징 정보 및 후보 콘텐츠들에 대한 제2특징 정보를 입력받는다(S610). 후보 콘텐츠들은 메타버스 환경에서 제공되는 콘텐츠들일 수 있다.

[0073] 그리고 컴퓨팅 장치는, 제1 및 제2특징 정보를 미리 강화 학습된 선호도 네트워크에 입력하여, 후보 콘텐츠들에 대한 사용자의 선호도값을 획득(S620)한다. 그리고 선호도값에 기반하여, 후보 콘텐츠들 중에서 추천 콘텐츠를 결정(S630)한다. 컴퓨팅 장치는 전술된 바와 같이, ε-탐욕(ε-greedy) 정책에 따라 적어도 하나의 추천 콘텐츠를 결정할 수 있다. 컴퓨팅 장치는 입실론(ε) 확률로 랜덤하게 추천 콘텐츠를 결정하고, 1-ε 확률로 선도도값이 최대인 추천 콘텐츠를 결정할 수 있다.

[0074] 추천 콘텐츠가 사용자에게 제공되면 컴퓨팅 장치는 추천 콘텐츠에 대한 사용자의 선호도 피드백 정보를 이용하여, 제1특징 정보 및 추천 콘텐츠에 대한 제3특징 정보를 갱신하고, 제1 및 제3특징 정보 및 갱신된 제1 및 제3특징 정보를 이용하여, 선호도 네트워크에 대한 강화 학습을 수행한다. 컴퓨팅 장치는 제1 및 제3특징 정보 사이의 내적값과, 선호도 피드백 정보에 포함된 보상값 사이의 차이에 따라서, 제1 및 제3특징 정보를 갱신할 수 있다.

[0075] 한편, 본 발명의 일실시예에 따른 컴퓨팅 장치는 추천 콘텐츠를 결정하여 사용자에게 제공할 뿐만 아니라, 메타버스 환경에서 추천 콘텐츠가 제공되는 경우, 사용자의 선호도 피드백 정보에 따라서 메타버스 환경을 변경할 수 있다. 즉, 추천 콘텐츠가 제공되는 메타버스 환경이, 선호도 피드백 정보에 따라 달라질 수 있다. 예컨대,

메타버스 환경이 미술관이라고 할 경우, 미술관의 디자인이나 색상, 형태가 사용자가 선호하는 디자인이나 색상 또는 형태로 변경될 수 있다.

[0077] 도 7은 본 발명의 구체적 실시예에 따른 강화 학습 기반의 콘텐츠 추천 방법을 설명하기 위한 도면으로서, 메타버스 미술관에서 미술작품 콘텐츠가 추천되는 방법이 일실시예로서 설명된다.

[0078] 미술작품의 속성은, 간색 계열, 녹색 계열, 파란색 계열, 검은색 계열, 흰색 계열의 5가지 속성으로 정의되며, 각 속성은 0에서 1사의 값으로 표현될 수 있다.

[0079] 사용자의 콘텐츠 선호도에 대한 제1특징 정보가, [0.25, 0.25, 0.25, 0.25, 0.25] 벡터로 정의되고, 추천 미술작품에 대한 제2특징 정보가 [0.4, 0.5, 0, 0.1, 0]로 정의되고, 보상값이 0.3인 경우, 제1특징 정보는 [수학식 1]에 따라 [0.249, 0.25, 0.249, 0.249, 0.249]로 갱신될 수 있다. 이 때, 상수값은 0.1이고, 학습률은 0.01이다.

[0080] 제2특징 정보 역시 [수학식 2]에 따라 갱신되며, 갱신된 제1 및 제2특징 정보에 의해 선호도 네트워크가 학습된다. 선호도 네트워크는 메타버스 미술관에 존재하는 후보 미술작품에 대한 선호도값을 출력하며, 선호도값에 따라 새로운 추천 미술작품이 결정되어 사용자에게 제공될 수 있다.

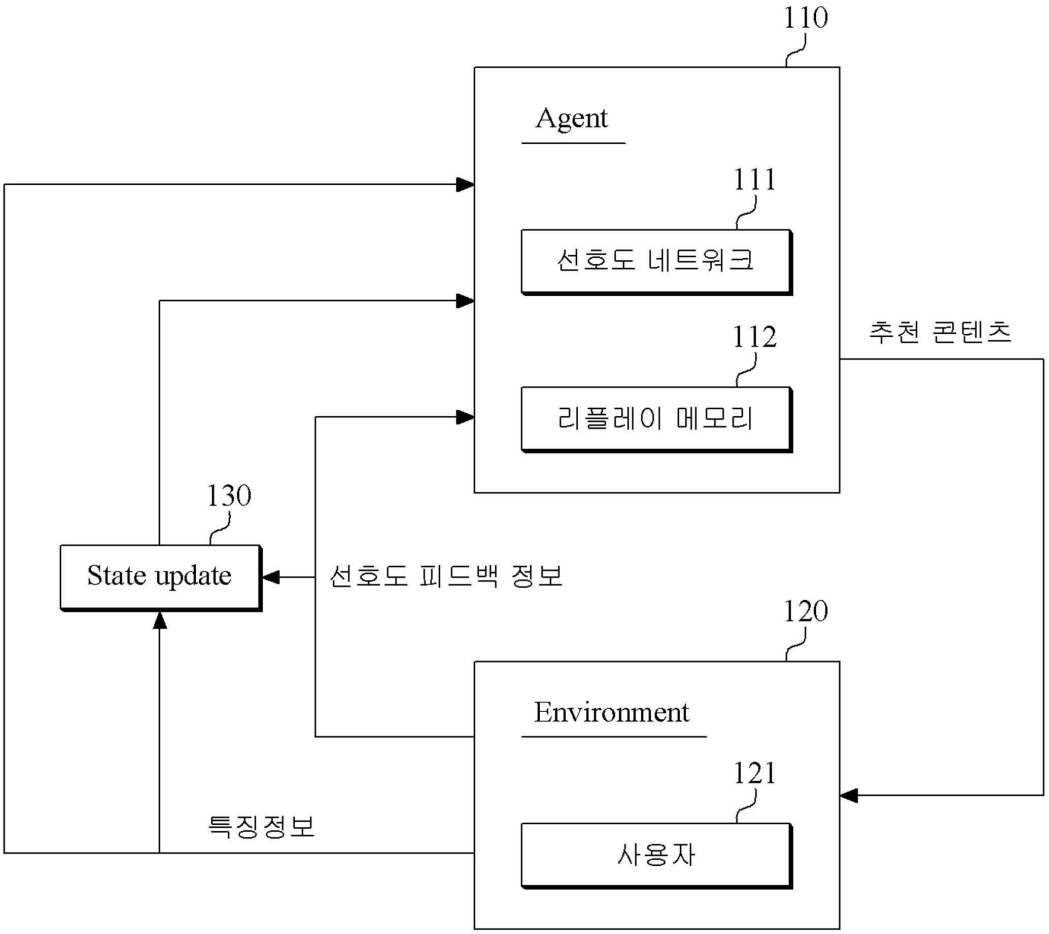
[0081] 추천 미술작품이 사용자에게 제공될 때, 메타버스 미술관의 구조나 디자인, 형태등이 추천 미술작품에 따라서 변경될 수 있다. 이와 같이, 메타버스 환경이 사용자에게 콘텐츠를 추천하고, 사용자는 추천된 콘텐츠에 대한 피드백을 제공하고, 다시 메타버스 환경은 피드백을 이용해 콘텐츠를 추천하거나 변화시킴으로서, 사용자와 메타버스 콘텐츠 사이의 상호 작용이 활발해지며, 이를 통해 - 지각적 어포던스 디자인 요소가 개선되고, 사용자의 경험 만족도가 개선될 수 있다.

[0083] 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.

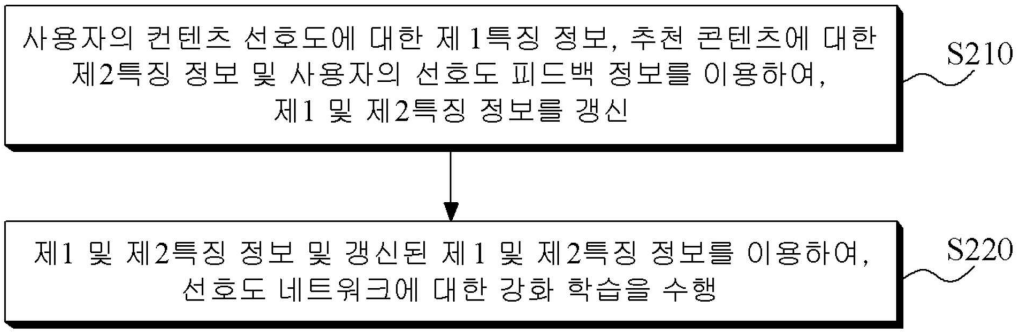
[0085] 이상과 같이 본 발명에서는 구체적인 구성 요소 등과 같은 특정 사항들과 한정된 실시예 및 도면에 의해 설명되었으나 이는 본 발명의 보다 전반적인 이해를 돕기 위해서 제공된 것일 뿐, 본 발명은 상기의 실시예에 한정되는 것은 아니며, 본 발명이 속하는 분야에서 통상적인 지식을 가진 자라면 이러한 기재로부터 다양한 수정 및 변형이 가능하다. 따라서, 본 발명의 사상은 설명된 실시예에 국한되어 정해져서는 아니되며, 후술하는 특허청구범위뿐 아니라 이 특허청구범위와 균등하거나 등가적 변형이 있는 모든 것들은 본 발명 사상의 범주에 속한다고 할 것이다.

도면

도면1


110
Agent
111
선호도 네트워크
112
리플레이 메모리
추천 콘텐츠
130
State update
선호도 피드백 정보
120
Environment
121
사용자
특징정보

도면2

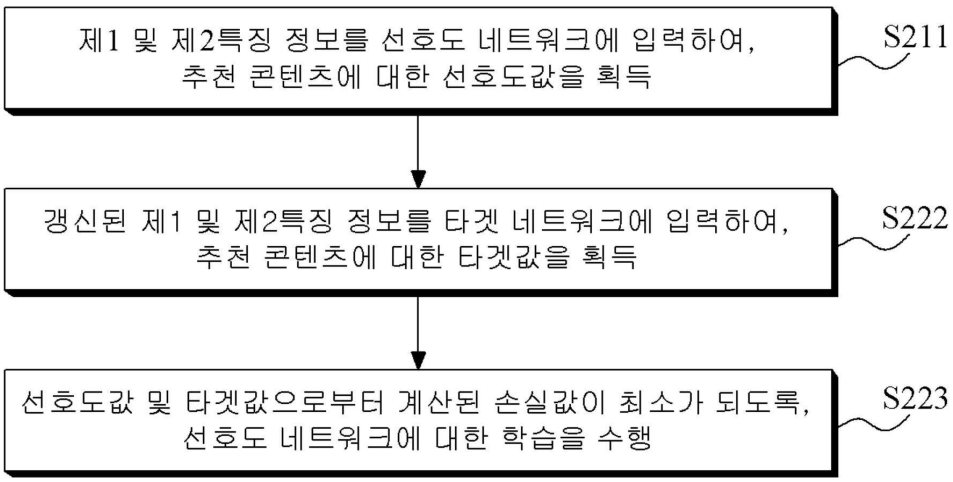

사용자의 컨텐츠 선호도에 대한 제1특징 정보, 추천 콘텐츠에 대한 제2특징 정보 및 사용자의 선호도 피드백 정보를 이용하여, 제1 및 제2특징 정보를 갱신
S210
제1 및 제2특징 정보 및 갱신된 제1 및 제2특징 정보를 이용하여, 선호도 네트워크에 대한 강화 학습을 수행
S220

도면3

Algorithm 1 Procedure of State update

Notations : the number of feature d, the number of content I, user state s_t, the set of content C, selected content index a_t, user feedback r, constant γ, learning rate α
State Update(t):
item i $\in \{1, ..., I\}$
if t = 0 **then**
    initialize $s_t \in [0,1]^d$ with prior information
    initialize $C_i \in [0,1]^d \subset C$ with prior information
else
    $s_{t+1} = s_t - 2\alpha[(s_t C_{a_t}^T - r)C_{a_t} + \gamma s_t]$
    $C_{a_t} = C_{a_t} - 2\alpha[(C_{a_t} s_t^T - r)s_t + \gamma C_{a_t}]$
end if
return s_{t+1}, C_{a_t}

도면4

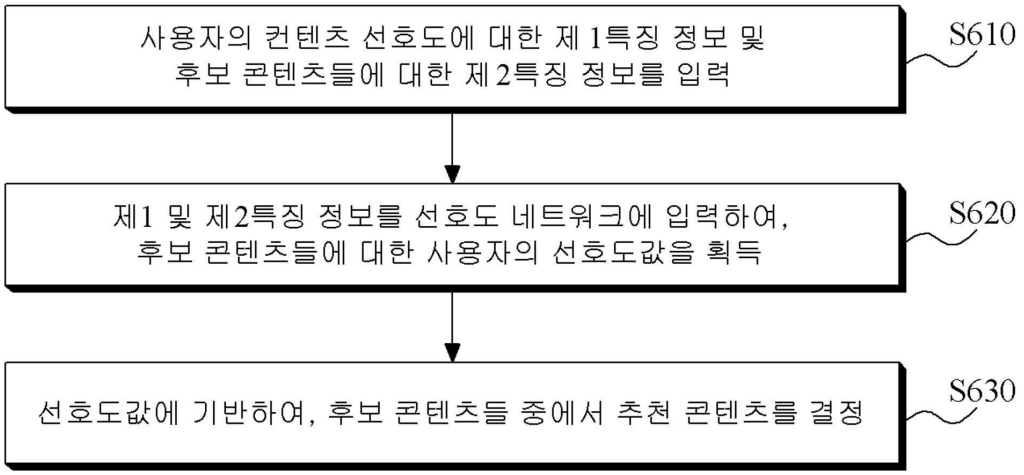


도면5

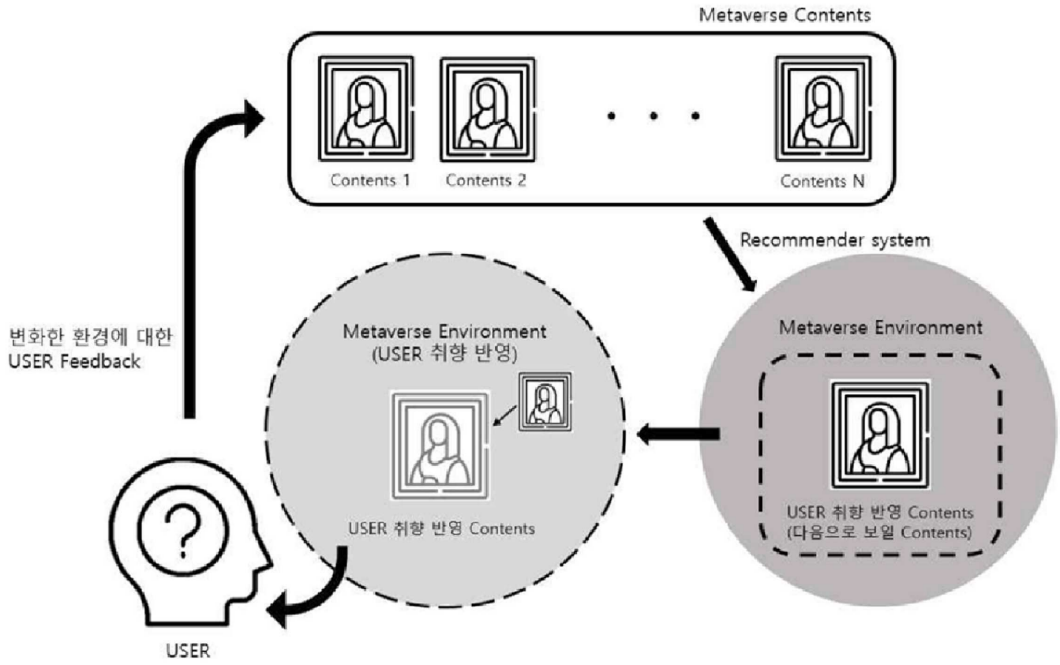
Algorithm 2 Procedure of updating the meta-verse environment change system

Notations : preference network Q with θ, target preference network $\hat{Q}$ with $\bar{\theta}$, user state s_t, the set of content information C, user feedback r, discount factor γ, epsilon ϵ, replay memory M, the number of contents I
Initialize $\hat{Q}$ with random weight , $\bar{\theta} = \theta$
for time-slot $t \in \{1, 2, ...\}$ **do**
    for $i \in \{1, 2, ...I\}$ **do**
        $Q(s_t, C_i; \theta)$
    end for
    With probability ϵ select a random action a_t
    otherwise select action a_t as the largest $Q(s_t, C_i; \theta)$
    Take a_t, observed feedback r
    $s_{t+1}, C_{a_t}^{t+1}$ = **State update** $(s_t, C_{a_t}^t, r)$
    if $i \neq a_t$ **then**
        $C_i^{t+1} \leftarrow C_i^t$, $\forall i \in \{1, 2, ...I\}$
    end if
    Store transition $(s_t, C_{a_t}^t, r, s_{t+1}, \{C_i^{t+1}\}_{\forall_{i \in \{1,2,...,I\}}})$
    Sample a mini-batch of $(s_j, C_{a_j}^j, r_j, s_{j+1}, \{C_i^{j+1}\}_{\forall_{i \in \{1,2,...,I\}}})$ from M
    $i^* \leftarrow \text{argmax}_{i \in I} \hat{Q}(s_{j+1}, C_i^{j+1}; \bar{\theta})$
    $y_j = r_j + \gamma \hat{Q}(s_{j+1}, C_{i^*}; \bar{\theta})$
    perform a gradient descent step on $(y_j - Q(s_j, C_{a_j}^j; \theta))^2$ with respect to the network parameter θ
    Update the target network $\bar{\theta} \leftarrow \theta$
end for

도면6


사용자의 컨텐츠 선호도에 대한 제1특징 정보 및
후보 콘텐츠들에 대한 제2특징 정보를 입력
S610
제1 및 제2특징 정보를 선호도 네트워크에 입력하여,
후보 콘텐츠들에 대한 사용자의 선호도값을 획득
S620
선호도값에 기반하여, 후보 콘텐츠들 중에서 추천 콘텐츠를 결정
S630

도면7


Metaverse Contents
Contents 1
Contents 2
Contents N
Recommender system
Metaverse Environment
USER 취향 반영 Contents
(다음으로 보일 Contents)
Metaverse Environment
(USER 취향 반영)
USER 취향 반영 Contents
변화한 환경에 대한
USER Feedback
USER