

(12) 특허협력조약에 의하여 공개된 국제출원

(19) 세계지식재산권기구
국제사무국

(43) 국제공개일
2021년 9월 16일 (16.09.2021) **WIPO | PCT**

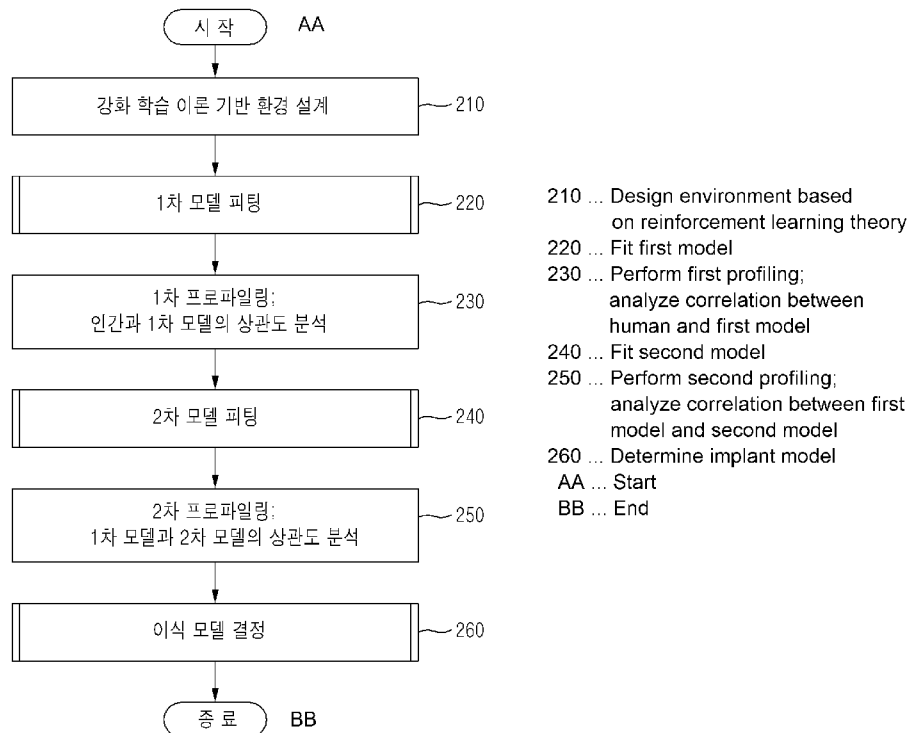


(10) 국제공개번호
WO 2021/182723 A1

- (51) 국제특허분류: *G06N 20/00* (2019.01) *G06N 3/00* (2006.01)
- (21) 국제출원번호: PCT/KR2020/018025
- (22) 국제출원일: 2020년 12월 10일 (10.12.2020)
- (25) 출원언어: 한국어
- (26) 공개언어: 한국어
- (30) 우선권정보:
10-2020-0028772 2020년 3월 9일 (09.03.2020) KR
10-2020-0126999 2020년 9월 29일 (29.09.2020) KR
- (71) 출원인: 한국과학기술원 (KOREA ADVANCED INSTITUTE OF SCIENCE AND TECHNOLOGY)
[KR/KR]: 34141 대전시 유성구 대학로 291, Daejeon (KR).
- (72) 발명자: 이상완 (LEE, Sang Wan); 34141 대전시 유성구 대학로 291, Daejeon (KR). 김동재 (KIM, Dongjae); 34141 대전시 유성구 대학로 291, Daejeon (KR). 신재훈 (SHIN, Jae Hoon); 34141 대전시 유성구 대학로 291, Daejeon (KR).
- (74) 대리인: 양성보 (YANG, Sungbo); 06099 서울시 강남구 선릉로125길 14 삼성빌딩 2층 피엔티특허법률사무소, Seoul (KR).
- (81) 지정국 (별도의 표시가 없는 한, 가능한 모든 종류의 국내 권리의 보호를 위하여): AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BN, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CZ, DE, DJ, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IR, IS, IT, JO, JP, KE, KG, KH, KN, KP, KW, KZ, LA, LC, LK, LR, LS, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PA, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, SA, SC, SD,

(54) Title: ELECTRONIC DEVICE FOR PRECISE BEHAVIORAL PROFILING FOR IMPLANTING HUMAN INTELLIGENCE INTO ARTIFICIAL INTELLIGENCE, AND OPERATION METHOD THEREFOR

(54) 발명의 명칭: 인간의 지능을 인공 지능으로 이식하기 위한 정밀 행동 프로파일링을 위한 전자 장치 및 그의 동작 방법



(57) Abstract: An electronic device and an operation method therefor according to various embodiments are for precise behavioral profiling for implanting human intelligence into artificial intelligence, and may be configured to: design theoretically at least one environmental variable; fit, on the basis of the environmental variable, a first model from human processing data about a task; fit, on the basis of the environmental variable, a second model from processing data of the first model about the task; and determine, via profiling for the first model and the second model, the second model as an implant model for human intelligence, on the basis of the correlation between the first model and the second model. According to various embodiments, the human processing data may comprise at least

SE, SG, SK, SL, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ,
UA, UG, US, UZ, VC, VN, WS, ZA, ZM, ZW.

- (84) 지정국 (별도의 표시가 없는 한, 가능한 모든 종류의 역
내 권리의 보호를 위하여): ARIPO (BW, GH, GM, KE,
LR, LS, MW, MZ, NA, RW, SD, SL, ST, SZ, TZ, UG, ZM,
ZW), 유라시아 (AM, AZ, BY, KG, KZ, RU, TJ, TM), 유
럽 (AL, AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI,
FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, MK,
MT, NL, NO, PL, PT, RO, RS, SE, SI, SK, SM, TR), OAPI
(BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, KM, ML,
MR, NE, SN, TD, TG).

공개:

— 국제조사보고서와 함께 (조약 제21조(3))

one of a brain signal or action data generated while a human processes the task.

(57) 요약서: 다양한 실시예들에 따른 전자 장치 및 그의 동작 방법은, 인간의 지능을 인공 지능으로 이식하기 위한 정밀
행동 프로파일링을 위한 것으로, 이론적으로 적어도 하나의 환경 변수를 설계하고, 환경 변수를 기반으로, 태스크에 대한
인간의 처리 데이터로부터 1차 모델을 피팅하고, 환경 변수를 기반으로, 태스크에 대한 1차 모델의 처리 데이터로부터 2차
모델을 피팅하고, 1차 모델과 2차 모델에 대한 프로파일링을 통해, 1차 모델과 2차 모델의 상관도를 기반으로, 2차 모델을
인간의 지능에 대한 이식 모델로 결정하도록 구성될 수 있다. 다양한 실시예들에 따르면, 인간의 처리 데이터는, 인간이
태스크를 처리하는 동안 발생하는 행위 데이터 또는 뇌 신호 중 적어도 어느 하나를 포함할 수 있다.

명세서

발명의 명칭: 인간의 지능을 인공 지능으로 이식하기 위한 정밀 행동 프로파일링을 위한 전자 장치 및 그의 동작 방법

기술분야

- [1] 다양한 실시예들은 인간의 지능을 인공 지능으로 이식하기 위한 정밀 행동 프로파일링을 위한 전자 장치 및 그의 동작 방법에 관한 것이다.

배경기술

- [2] 기존 인간 지능에 대한 연구는 그에 수반되는 의사결정 과정이 내재적이며 은닉되어 있기에, 그 과정을 모방하는 모델기반의 분석이 주된 연구 방법론이다. 이 방법론에서는 최대 우도(maximum likelihood)로 인간의 행동을 설명하기 위한 최적 모델을 선정하고 그 모델을 기반으로 뇌 내에서 이루어지는 인간 지능을 설명한다. 그러나, 이러한 과정은 최적 모델의 선정 기준이 실제 작업 수행에 필요한 특성과 독립적이고, 내재적으로 존재하는 과적합(overfitting)의 위험을 판단할 수 없으며, 특히 과적합의 위험성이 높은 심층신경망 기반의 인공 지능으로의 이식이 불가능하다는 한계가 있다.

발명의 상세한 설명

기술적 과제

- [3] 다양한 실시예들은, 인간의 행동 프로파일을 예측하는 인공 지능을 개발하기 위한 전자 장치 및 그의 동작 방법을 제공한다.
- [4] 다양한 실시예들은, 인간의 지능을 인공 지능으로 이식하기 위한 정밀 행동 프로파일링을 위한 전자 장치 및 그의 동작 방법을 제공한다.

기술적 해결방법

- [5] 다양한 실시예들에 따른 전자 장치의 동작 방법은, 태스크에 대한 인간의 처리 데이터를 기반으로, 1차 모델을 피팅하는 동작, 상기 태스크에 대한 상기 1차 모델의 처리 데이터를 기반으로, 2차 모델을 피팅하는 동작, 및 상기 1차 모델과 상기 2차 모델에 대한 프로파일링을 통해, 상기 2차 모델을 상기 인간의 지능에 대한 이식 모델로 결정하는 동작을 포함할 수 있다.
- [6] 다양한 실시예들에 따른 전자 장치는, 메모리, 및 상기 메모리와 연결되고, 상기 메모리에 저장된 적어도 하나의 명령을 실행하도록 구성된 프로세서를 포함하고, 상기 프로세서는, 태스크에 대한 인간의 처리 데이터를 기반으로, 1차 모델을 피팅하고, 상기 태스크에 대한 상기 1차 모델의 처리 데이터를 기반으로, 2차 모델을 피팅하고, 상기 1차 모델과 상기 2차 모델에 대한 프로파일링을 통해, 상기 2차 모델을 상기 인간의 지능에 대한 이식 모델로 결정하도록 구성될 수 있다.
- [7] 다양한 실시예들에 따른 컴퓨터 프로그램은, 컴퓨터 장치에 결합되어, 상기 컴퓨터 장치에 의해 판독 가능한 기록 매체에 저장되며, 태스크에 대한 인간의

처리 데이터를 기반으로, 1차 모델을 피팅하는 동작, 상기 태스크에 대한 상기 1차 모델의 처리 데이터를 기반으로, 2차 모델을 피팅하는 동작, 및 상기 1차 모델과 상기 2차 모델에 대한 프로파일링을 통해, 상기 2차 모델을 상기 인간의 지능에 대한 이식 모델로 결정하는 동작을 실행할 수 있다.

발명의 효과

- [8] 다양한 실시예들에 따르면, 인간의 지능과 유사한 인공 지능이 개발될 수 있다. 인간의 지능에 대한 고위수준 지표인 정밀 행동프로파일을 모사할 수 있는 이식 모델이 개발되고, 과적합의 위험 없이, 이식 모델이 인공 지능으로 이식될 수 있다. 이에 따라, 인공 지능이 인간의 행동 프로파일을 복원 가능하여, IoT 분야를 포함하는 인공 지능 비서와 같은 인간 보조 시스템 전반에 인간의 행동 범주 내에서 인간의 행동을 이해 및 예측할 수 있다.

도면의 간단한 설명

- [9] 도 1은 다양한 실시예들에 따른 전자 장치를 도시하는 도면이다.
 [10] 도 2는 다양한 실시예들에 따른 전자 장치의 동작 방법을 도시하는 도면이다.
 [11] 도 3a 및 도 3b는 도 2의 강화 학습 이론 기반 환경 설계 동작을 설명하기 위한 도면들이다.
 [12] 도 4는 도 2의 1차 모델 피팅 동작을 도시하는 도면이다.
 [13] 도 5는 도 2의 1차 모델 피팅 동작을 설명하기 위한 도면이다.
 [14] 도 6은 도 2의 2차 모델 피팅 동작을 도시하는 도면이다.
 [15] 도 7은 도 2의 2차 모델 피팅 동작을 설명하기 위한 도면이다.
 [16] 도 8 및 도 9는 도 2의 2차 프로파일링 동작을 설명하기 위한 도면들이다.
 [17] 도 10은 도 2의 이식 모델 결정 동작을 도시하는 도면이다.
 [18] 도 11은 도 2의 이식 모델 결정 동작을 설명하기 위한 도면이다.
 [19] 도 12는 다양한 실시예들에 따른 일반화 가능한 인간 모사형 강화학습 모델 설계를 위한 정량화 방법을 나타내는 흐름도이다.
 [20] 도 13은 다양한 실시예들에 따른 일반화 가능한 인간 모사형 강화학습 모델 설계를 위한 정량화 장치를 개략적으로 나타내는 블록도이다.
 [21] 도 14는 다양한 실시예들에 따른 인간의 잠재 정책 학습, 신뢰도 시험 및 경험적 일반화 시험을 설명하기 위한 도면이다.
 [22] 도 15는 다양한 실시예들에 따른 실험에 사용된 RL 모델의 구조를 설명하기 위한 도면이다.
 [23] 도 16은 다양한 실시예들에 따른 각 RL 모델의 일반화 시험을 위한 시뮬레이션 환경을 설명하기 위한 도면이다.
 [24] 도 17은 다양한 실시예들에 따른 RL 모델의 적응 능력에 대한 시뮬레이션 결과를 나타내는 도면이다.

발명의 실시를 위한 형태

- [25] 이하, 본 문서의 다양한 실시예들이 첨부된 도면을 참조하여 설명된다.

[26]

[27] 다양한 실시예들은, 인간의 지능을 인공 지능으로 이식하기 위한 정밀 행동 프로파일링을 위한 전자 장치 및 그의 동작 방법을 제공한다.

[28]

다양한 실시예들에 따르면, 인간의 작업 수행 특성과 동일한 특성을 갖는 모델이 제공된다. (1) 인간 작업 수행 과정에 대한 정밀 프로파일링 통한 모델의 개발: 인간 작업 수행 특성을 분석하고 이를 통해 계산 모델을 개발, 실제 작업 수행에 필요한 특성을 복원하는 모델을 개발할 수 있다. (2) 정밀 행동프로파일 비교를 통한 과적합(overfitting) 여부 판단: 실제 행동 프로파일과 (1)에서 개발한 모델의 행동 프로파일 간 비교를 통해 과적합의 여부를 평가할 수 있다. (3) 인간 지능-인공 지능 이식: 인간 지능에 대한 고위수준 지표인 정밀 행동프로파일을 모사할 수 있는 모델을 통해 과적합의 위험 없이 인공 지능으로 이식할 수 있다.

[29]

다양한 실시예들에 따르면, 인간 작업수행 과정 정밀 프로파일링과, 그에 기반한 인간 지능 모델 개발 기술, 그리고 과적합의 위험을 제거하여 인공 지능으로 이식하는 기술이 제공된다. 이러한 작업 수행 특성 정밀 프로파일링을 기반으로 한 인간 지능-인공 지능 이식 기술은 뇌 인지 기반 및 뇌 모사형 인공 지능 개발에 있어서 핵심적인 기술이며, 기존에 유사한 연구 사례가 없는 기술이다.

[30]

구체적으로, (1) 인간 작업 수행 특성을 통한 모델의 개발은 인간의 실제 행동으로부터 환경 변화에 대응하여 변화하는 작업 수행 특성을 행동 프로파일로 추출, 이에 따라 후보 모델을 만들고 서로 비교하여 최적 모델을 선정한다. 선정된 최적 모델은 작업 수행에 대한 인간의 행동 프로파일을 그대로 복원한다. (2) 행동 프로파일 비교를 통한 과적합 여부 판단은 (1)에서 선정한 최적 모델이 작업을 수행하면서 보이는 행동을 통하여 모델의 프로파일을 재차 추출하고, 이를 실제 행동 프로파일과 비교한다. 두 행동 프로파일간 비교는 정성적.정량적으로 이루어 질 수 있으며 정성적으로는 단순히 두 프로파일의 경향을, 정량적으로는 행동에 영향을 끼치는 핵심 파라미터의 분포에 대한 상관관계 분석 등을 통해 비교할 수 있다. (3) 인간 지능-인공 지능 이식은 (2)에서 이루어진 정성적.정량적 판단 기준에 따라 과적합의 위험 없이 심층신경망 기반으로 이루어질 수 있다.

[31]

다양한 실시예들은, (1) 작업수행 과정에 대한 정밀 행동 프로파일링과 이를 통한인간 지능 모델 개발, (2) 행동 프로파일링을 통한 인간 지능 모델의 과적합 여부 검정을 목적으로 하고 있으며, 궁극적으로 이를 통해 (3) 인간 지능-인공 지능의 이식에 있어서 과적합의 위험 없이 심층 신경망 기반 인공지능으로 이식할 수 있다.

[32]

[33]

도 1은 다양한 실시예들에 따른 전자 장치(100)를 도시하는 도면이다.

[34]

도 1을 참조하면, 다양한 실시예들에 따른 전자 장치(100)는 입력 모듈(110), 출력 모듈(120), 메모리(130) 또는 프로세서(140) 중 적어도 어느 하나를 포함할

수 있다. 어떤 실시예에서, 전자 장치(100)의 구성 요소들 중 적어도 어느 하나가 생략될 수 있으며, 적어도 하나의 다른 구성 요소가 추가될 수 있다. 어떤 실시예에서, 전자 장치(100)의 구성 요소들 중 적어도 어느 두 개가 하나의 통합된 회로로 구현될 수 있다.

- [35] 입력 모듈(110)은 전자 장치(100)의 적어도 하나의 구성 요소에 사용될 신호를 입력할 수 있다. 입력 모듈(110)은, 사용자가 전자 장치(100)에 직접적으로 신호를 입력하도록 구성되는 입력 장치, 주변의 변화를 감지하여 신호를 발생하도록 구성되는 센서 장치 또는 외부 기기로부터 신호를 수신하도록 구성되는 수신 장치 중 적어도 어느 하나를 포함할 수 있다. 예를 들면, 입력 장치는 마이크로폰(microphone), 마우스(mouse) 또는 키보드(keyboard) 중 적어도 어느 하나를 포함할 수 있다. 어떤 실시예에서, 입력 장치는 터치를 감지하도록 설정된 터치 회로(touch circuitry) 또는 터치에 의해 발생하는 힘의 세기를 측정하도록 설정된 센서 회로 중 적어도 어느 하나를 포함할 수 있다.
- [36] 출력 모듈(120)은 전자 장치(100)의 외부로 정보를 출력할 수 있다. 출력 모듈(120)은, 정보를 시각적으로 출력하도록 구성되는 표시 장치, 정보를 오디오 신호로 출력할 수 있는 오디오 출력 장치 또는 정보를 무선으로 송신할 수 있는 송신 장치 중 적어도 어느 하나를 포함할 수 있다. 예를 들면, 표시 장치는 디스플레이, 홀로그램 장치 또는 프로젝터 중 적어도 어느 하나를 포함할 수 있다. 일 예로, 표시 장치는 입력 모듈(110)의 터치 회로 또는 센서 회로 중 적어도 어느 하나와 조립되어, 터치 스크린으로 구현될 수 있다. 예를 들면, 오디오 출력 장치는 스피커 또는 리시버 중 적어도 어느 하나를 포함할 수 있다.
- [37] 일 실시예에 따르면, 수신 장치와 송신 장치는 통신 모듈로 구현될 수 있다. 통신 모듈은 전자 장치(100)에서 외부 기기와 통신을 수행할 수 있다. 통신 모듈은 전자 장치(100)와 외부 기기 간 통신 채널을 수립하고, 통신 채널을 통해, 외부 기기와 통신을 수행할 수 있다. 여기서, 외부 기기는 위성, 기지국, 서버 또는 다른 전자 장치 중 적어도 어느 하나를 포함할 수 있다. 통신 모듈은 유선 통신 모듈 또는 무선 통신 모듈 중 적어도 어느 하나를 포함할 수 있다. 유선 통신 모듈은 외부 기기와 유선으로 연결되어, 유선으로 통신할 수 있다. 무선 통신 모듈은 근거리 통신 모듈 또는 원거리 통신 모듈 중 적어도 어느 하나를 포함할 수 있다. 근거리 통신 모듈은 외부 기기와 근거리 통신 방식으로 통신할 수 있다. 예를 들면, 근거리 통신 방식은, 블루투스(Bluetooth), 와이파이 다이렉트(WiFi direct), 또는 적외선 통신(IrDA; infrared data association) 중 적어도 어느 하나를 포함할 수 있다. 원거리 통신 모듈은 외부 기기와 원거리 통신 방식으로 통신할 수 있다. 여기서, 원거리 통신 모듈은 네트워크를 통해 외부 기기와 통신할 수 있다. 예를 들면, 네트워크는 셀룰러 네트워크, 인터넷, 또는 LAN(local area network)이나 WAN(wide area network)과 같은 컴퓨터 네트워크 중 적어도 어느 하나를 포함할 수 있다.
- [38] 메모리(130)는 전자 장치(100)의 적어도 하나의 구성 요소에 의해 사용되는

다양한 데이터를 저장할 수 있다. 예를 들면, 메모리(130)는 휘발성 메모리 또는 비휘발성 메모리 중 적어도 어느 하나를 포함할 수 있다. 데이터는 적어도 하나의 프로그램 및 이와 관련된 입력 데이터 또는 출력 데이터를 포함할 수 있다. 프로그램은 메모리(130)에 적어도 하나의 명령을 포함하는 소프트웨어로서 저장될 수 있으며, 운영 체제, 미들웨어 또는 어플리케이션 중 적어도 어느 하나를 포함할 수 있다.

[39] 프로세서(140)는 메모리(130)의 프로그램을 실행하여, 전자 장치(100)의 적어도 하나의 구성 요소를 제어할 수 있다. 이를 통해, 프로세서(140)는 데이터 처리 또는 연산을 수행할 수 있다. 이 때 프로세서(140)는 메모리(130)에 저장된 명령을 실행할 수 있다.

[40] 프로세서(140)는 인간의 지능을 인공 지능으로 이식하기 위한 강화 학습 이론 기반 환경을 설계할 수 있다. 이 때 프로세서(140)는 인간의 태스크 처리와 관련된 환경을 설계할 수 있다. 여기서, 프로세서(140)는, 예컨대 벨만 방정식(Bellman equation)에 기반하여 적어도 하나의 환경 변수(environmental factor)를 결정하고, 그 값을 최적화할 수 있다. 예를 들면, 환경 변수는 상태-전이 불확실성(state-transition uncertainty), 상태-공간 복잡성(state-space complexity), 신규성(novelty), 상태 예측 오류(state prediction error) 또는 보상 예측 오류(reward prediction error) 중 적어도 어느 하나를 포함할 수 있다.

[41] 프로세서(140)는 환경 변수를 기반으로, 1차 모델(first level model)을 피팅(fitting)할 수 있다. 프로세서(140)는 환경 변수를 기반으로, 태스크에 대한 인간의 처리 데이터로부터 1차 모델을 피팅할 수 있다. 이 때 태스크에 대한 인간의 처리 데이터는, 인간이 태스크를 처리하는 동안 발생하는 행위 데이터 또는 뇌 신호 중 적어도 어느 하나를 포함할 수 있다. 그리고, 프로세서(140)는 인간과 1차 모델에 대한 프로파일링, 즉 1차 프로파일링을 수행할 수 있다. 이를 통해, 프로세서(140)는 인간과 1차 모델의 상관도를 분석할 수 있다. 예를 들면, 상관도는 최대 1이며, 인간과 1차 모델이 동일할 때, 상관도는 1일 수 있다. 이 때 프로세서(140)는 태스크에 대한 인간의 처리 데이터에 대해, 1차 모델의 과적합(overfitting) 여부를 판단할 수 있다. 이를 위해, 프로세서(140)는, 인간이 태스크를 처리하는 데 있어서의 행동 프로파일과 1차 모델의 행동 프로파일을 비교할 수 있다. 한편, 프로세서(140)는, 인간이 태스크를 처리하는 데 있어서의 파라미터와 1차 모델의 파라미터를 비교할 수 있다.

[42] 프로세서(140)는 2차 모델(second level model)을 피팅할 수 있다. 프로세서(140)는 환경 변수를 기반으로, 태스크에 대한 1차 모델의 처리 데이터로부터 2차 모델을 피팅할 수 있다. 그리고, 프로세서(140)는 2차 프로파일을 수행할 수 있다. 이를 통해, 전자 장치(100)는 1차 모델과 2차 모델의 상관도를 분석할 수 있다. 이 때 프로세서(140)는 1차 모델의 행동 프로파일과 2차 모델의 행동 프로파일을 비교할 수 있다. 한편, 프로세서(140)는 1차 모델의 파라미터와 2차 모델의 파라미터를 비교할 수 있다. 이를 통해, 프로세서(140)가

1차 모델과 2차 모델의 상관도를 검출할 수 있다.

[43] 프로세서(140)는 인간 지능에 대한 이식 모델을 결정할 수 있다.

프로세서(140)는 1차 모델과 2차 모델의 상관도를 기반으로, 2차 모델을 이식 모델로 결정할 수 있다. 이 때 1차 모델과 2차 모델의 상관도는 1차 모델과 2차 모델의 유사한 정도를 나타낼 수 있다. 이를 통해, 1차 모델과 2차 모델이 일정 수준 이상으로 유사하면, 프로세서(140)는 2차 모델을 이식 모델로 결정할 수 있다. 예를 들면, 상관도는 최대 1이며, 1차 모델과 2차 모델이 동일할 때, 상관도는 1일 수 있다.

[44]

[45] 도 2는 다양한 실시예들에 따른 전자 장치(100)의 동작 방법을 도시하는 도면이다. 그리고, 도 3a, 도 3b, 도 4, 도 5, 도 6, 도 7, 도 8, 도 9, 도 10 및 도 11은 다양한 실시예들에 따른 전자 장치(100)의 동작 방법을 예시적으로 설명하기 위한 도면들이다.

[46] 도 2를 참조하면, 전자 장치(100)는 210 동작에서 인간의 지능을 인공 지능으로 이식하기 위한 강화 학습 이론 기반 환경을 설계할 수 있다. 이 때 프로세서(140)는 인간의 태스크 처리와 관련된 환경을 설계할 수 있다. 예를 들면, 프로세서(140)는, 인간이 태스크를 처리하는 데 있어서의 작업 수행 과정 또는 문제 해결 과정 중 적어도 어느 하나를 설명할 수 있는 강화학습 이론을 바탕으로, 인간을 위한 표준 작업 환경을 설계할 수 있다. 여기서, 프로세서(140)는, 예컨대 벨만 방정식(Bellman equation)에 기반하여 적어도 하나의 환경 변수(environmental factor)를 결정하고, 그 값을 최적화할 수 있다. 예를 들면, 환경 변수는 상태-전이 불확실성(state-transition uncertainty), 상태-공간 복잡성(state-space complexity), 신규성(novelty), 상태 예측 오류(state prediction error) 또는 보상 예측 오류(reward prediction error) 중 적어도 어느 하나를 포함할 수 있다. 이에 대해, 도 3a 및 도 3b를 참조하여, 보다 상세하게 후술될 것이다.

[47] 도 3a 및 도 3b는 도 2의 강화 학습 이론 기반 환경 설계 동작(210)을 설명하기 위한 도면들이다.

[48] 도 3a를 참조하면, 강화 학습 이론 기반 환경은, 인간이 태스크를 처리하는 데 있어서 발생 가능한 적어도 하나의 상태(state), 각 상태에서 인간에 의해 이루어지는 적어도 하나의 의사 결정(choice) 및 각 의사 결정에 따른 적어도 하나의 상태 전이(state-transition)로 표현될 수 있다. 이 때 각 노드가 각 상태를 나타내고, 각 화살표가 각 의사 결정을 나타내며, 각 실선이 각 상태 전이를 나타낼 수 있다. 도 3b에 도시된 바와 같이, 일 상태(S_t)에서의 의사 결정을 기반으로, 다른 상태(S_{t+1})로의 상태 전이가 이루어질 수 있다. 각 상태 전이는, 상태-전이 가능성(state-transition probability)을 가질 수 있다. 예를 들면, 각 상태에 대해 복수 개의 의사 결정들이 가능하므로, 도 3b에 도시된 바와 같이 상태-공간 복잡성이 정의될 수 있다. 여기서, 각 상태에 대해 의사 결정들의

개수가 많을수록, 상태-공간 복잡성이 높을 수 있다. 예를 들면, 각 의사 결정에 대해 복수 개의 상태 전이들이 가능하므로, 도 3b에 도시된 바와 같이, 상태-전이 불확실성이 정의될 수 있다. 여기서, 각 의사 결정에 대한 상태 전이들의 가능성들 간 차이값이 클수록, 상태-전이 불확실성이 낮을 수 있다.

[49]

[50] 다시 도 2를 참조하면, 전자 장치(100)는 220 동작에서 환경 변수를 기반으로, 1차 모델(first level model)을 피팅(fitting)할 수 있다. 프로세서(140)는 환경 변수를 기반으로, 태스크에 대한 인간의 처리 데이터로부터 1차 모델을 피팅할 수 있다. 이 때 태스크에 대한 인간의 처리 데이터는, 인간이 태스크를 처리하는 동안 발생하는 행위 데이터 또는 뇌 신호 중 적어도 어느 하나를 포함할 수 있다. 이에 대해, 도 4 및 도 5를 참조하여, 보다 상세하게 후술될 것이다.

[51] 도 4는 도 2의 1차 모델 피팅 동작(220)을 도시하는 도면이다. 그리고, 도 5는 도 2의 1차 모델 피팅 동작(220)을 설명하기 위한 도면이다.

[52] 도 4를 참조하면, 전자 장치(100)는 410 동작에서 태스크에 대한 인간의 처리 데이터를 수집할 수 있다. 프로세서(140)는, 인간이 실질적으로 태스크를 처리하는 과정을 추적하면서, 태스크에 대한 인간의 처리 데이터를 수집할 수 있다. 여기서, 프로세서(140)는 입력 모듈(110)을 통해 인간의 처리 데이터를 수집할 수 있다. 예를 들면, 프로세서(140)는 입력 장치 또는 통신 모듈을 통해 인간의 행위 데이터(behavioral data)를 수집하고, 센서 장치를 통해 인간의 뇌 신호를 수집할 수 있다. 일 예로, 뇌 신호는 기능적 MRI(FMRI; functional magnetic resonance imaging) 신호를 포함할 수 있다.

[53] 전자 장치(100)는 420 동작에서 태스크에 대한 인간의 처리 데이터를 기반으로, 1차 모델을 학습할 수 있다. 프로세서(140)는 환경 변수를 기반으로, 태스크에 대한 인간의 처리 데이터로부터 1차 모델을 학습할 수 있다. 이 때 1차 모델의 행동 프로파일 또는 적어도 하나의 파라미터 중 적어도 어느 하나가 검출될 수 있다. 예를 들면, 프로세서(140)는, 도 5의 (a)에 도시된 바와 같이 1차 모델의 행동 프로파일을 검출할 수 있다. 여기서, 1차 모델의 행동 프로파일은 상태-공간 복잡성 또는 상태-전이 불확실성 중 적어도 어느 하나로부터 검출될 수 있다. 예를 들면, 프로세서(140)는, 도 5의 (b)에 도시된 바와 같이 1차 모델의 파라미터를 검출할 수 있다. 여기서, 1차 모델의 파라미터는 상태-전이 불확실성, 상태-공간 복잡성, 이전의 상태로부터의 상태 전이에 따른 보상(reward), 이전의 상태로부터의 상태 전이에 따른 반응(action) 또는 최대 목표값 중 적어도 어느 하나를 포함할 수 있다. 이후 전자 장치(100)는 도 2로 리턴하여, 230 동작으로 진행할 수 있다.

[54]

[55] 다시 도 2를 참조하면, 전자 장치(100)는 230 동작에서 인간과 1차 모델에 대한 프로파일링, 즉 1차 프로파일링을 수행할 수 있다. 이를 통해, 전자 장치(100)는 인간과 1차 모델의 상관도를 분석할 수 있다. 예를 들면, 상관도는 최대 1이며,

인간과 1차 모델이 동일할 때, 상관도는 1일 수 있다. 이 때 프로세서(140)는 태스크에 대한 인간의 처리 데이터에 대해, 1차 모델의 과적합(overfitting) 여부를 판단할 수 있다. 이를 위해, 프로세서(140)는, 인간이 태스크를 처리하는 데 있어서의 행동 프로파일과 1차 모델의 행동 프로파일을 비교할 수 있다. 한편, 프로세서(140)는, 인간이 태스크를 처리하는 데 있어서의 파라미터와 1차 모델의 파라미터를 비교할 수 있다.

- [56] 전자 장치(100)는 240 동작에서 2차 모델(second level model)을 피팅할 수 있다. 프로세서(140)는 환경 변수를 기반으로, 태스크에 대한 1차 모델의 처리 데이터로부터 2차 모델을 피팅할 수 있다. 이에 대해, 도 6 및 도 7을 참조하여, 보다 상세하게 후술될 것이다.
- [57] 도 6은 도 2의 2차 모델 피팅 동작(240)을 도시하는 도면이다. 그리고, 도 7은 도 2의 2차 모델 피팅 동작(240)을 설명하기 위한 도면이다.
- [58] 도 6을 참조하면, 전자 장치(100)는 610 동작에서 태스크에 대한 1차 모델의 처리 데이터를 수집할 수 있다. 프로세서(140)는, 1차 모델이 태스크를 처리하는 과정을 추적하면서, 태스크에 대한 1차 모델의 처리 데이터를 수집할 수 있다. 이 때 프로세서(140)는, 410 동작에서 인간에 의해 수행된 태스크를 1차 모델을 이용하여 재차 처리하고, 이를 통해 태스크에 대한 1차 모델의 처리 데이터를 수집할 수 있다.
- [59] 전자 장치(100)는 620 동작에서 태스크에 대한 1차 모델의 처리 데이터를 기반으로, 2차 모델을 학습할 수 있다. 프로세서(140)는 환경 변수를 기반으로, 태스크에 대한 1차 모델의 처리 데이터로부터 2차 모델을 학습할 수 있다. 이 때 2차 모델의 행동 프로파일 또는 적어도 하나의 파라미터 중 적어도 어느 하나가 검출될 수 있다. 예를 들면, 프로세서(140)는, 도 7의 (a)에 도시된 바와 같이 2차 모델의 행동 프로파일을 검출할 수 있다. 여기서, 2차 모델의 행동 프로파일은 상태-공간 복잡성 또는 상태-전이 불확실성 중 적어도 어느 하나로부터 검출될 수 있다. 예를 들면, 프로세서(140)는, 도 7의 (b)에 도시된 바와 같이 2차 모델의 파라미터를 검출할 수 있다. 여기서, 2차 모델의 파라미터는 상태-전이 불확실성, 상태-공간 복잡성, 이전의 상태로부터의 상태 전이에 따른 보상, 이전의 상태로부터의 상태 전이에 따른 반응 또는 최대 목표값 중 적어도 어느 하나를 포함할 수 있다. 이 후 전자 장치(100)는 도 2로 리턴하여, 250 동작으로 진행할 수 있다.
- [60]
- [61] 다시 도 2를 참조하면, 전자 장치(100)는 250 동작에서 2차 프로파일을 수행할 수 있다. 이를 통해, 전자 장치(100)는 1차 모델과 2차 모델의 상관도를 분석할 수 있다. 이 때 프로세서(140)는 1차 모델의 행동 프로파일과 2차 모델의 행동 프로파일을 비교할 수 있다. 한편, 프로세서(140)는 1차 모델의 파라미터와 2차 모델의 파라미터를 비교할 수 있다. 이를 통해, 프로세서(140)가 1차 모델과 2차 모델의 상관도를 검출할 수 있다. 이에 대해, 도 8 및 도 9를 참조하여, 보다

상세하게 후술될 것이다.

[62] 도 8 및 도 9는 도 2의 2차 프로파일링 동작(250)을 설명하기 위한 도면들이다.

[63] 도 8 및 도 9를 참조하면, 프로세서(140)는 1차 모델과 2차 모델을 비교하여, 1차 모델과 2차 모델의 상관도를 검출할 수 있다. 이를 위해, 프로세서(140)는, 도 8의 (a)에 도시된 바와 같은 1차 모델의 행동 프로파일과 도 8의 (b)에 도시된 바와 같은 2차 모델의 행동 프로파일을 정성적으로(qualitatively) 비교할 수 있다. 여기서, 프로세서(140)는 1차 모델의 행동 프로파일과 2차 모델의 행동 프로파일을 비교하여, 프로파일 상관도를 검출할 수 있다. 한편, 프로세서(140)는, 도 9의 (a)와 (b)에 각각 도시된 바와 같이 1차 모델의 파라미터와 2차 모델의 파라미터를 정량적으로(quantitatively) 각각 비교할 수 있다. 여기서, 프로세서(140)는 1차 모델의 파라미터와 2차 모델의 파라미터를 비교하여, 파라미터 상관도를 검출할 수 있다. 그리고, 프로세서(140)는 프로파일 상관도 또는 파라미터 상관도 중 적어도 어느 하나를 기반으로, 1차 모델과 2차 모델의 상관도를 검출할 수 있다.

[64]

[65] 다시 도 2를 참조하면, 전자 장치(100)는 260 동작에서 인간 지능에 대한 이식 모델을 결정할 수 있다. 프로세서(140)는 1차 모델과 2차 모델의 상관도를 기반으로, 2차 모델을 이식 모델로 결정할 수 있다. 이 때 1차 모델과 2차 모델의 상관도는 1차 모델과 2차 모델의 유사한 정도를 나타낼 수 있다. 이를 통해, 1차 모델과 2차 모델이 일정 수준 이상으로 유사하면, 프로세서(140)는 2차 모델을 이식 모델로 결정할 수 있다. 예를 들면, 상관도는 최대 1이며, 1차 모델과 2차 모델이 동일할 때, 상관도는 1일 수 있다. 이에 대해, 도 10 및 도 11을 참조하여, 보다 상세하게 후술될 것이다.

[66] 도 10은 도 2의 이식 모델 결정 동작(260)을 도시하는 도면이다. 그리고, 도 11은 도 2의 이식 모델 결정 동작(260)을 설명하기 위한 도면이다.

[67] 도 10을 참조하면, 전자 장치(100)는 1010 동작에서 1차 모델과 2차 모델의 상관도를 미리 설정되는 임계값과 비교할 수 있다. 프로세서(140)는, 1차 모델과 2차 모델의 상관도가 1이하이면서, 임계값을 초과하는지의 여부를 판단할 수 있다. 일 예로, 1차 모델과 2차 모델의 상관도가 높으면, 1차 모델과 2차 모델은, 도 11의 (a)에 도시된 바와 같은 관계를 나타낼 수 있다. 다른 예로, 1차 모델과 2차 모델의 상관도가 낮으면, 1차 모델과 2차 모델은, 도 11의 (b)에 도시된 바와 같은 관계를 나타낼 수 있다.

[68] 1010 동작에서 1차 모델과 2차 모델의 상관도가 임계값 이하인 것으로 판단되면, 전자 장치(100)는 도 2로 리턴하여, 220 동작으로 복귀할 수 있다. 즉 1차 모델과 2차 모델이 일정 수준 미만으로 상이하면, 프로세서(140)는 2차 모델을 이식 모델로 결정하지 않고, 220 동작으로 복귀할 수 있다. 그리고, 프로세서(140)는 220 동작 내지 260 동작을 반복하여 수행할 수 있다.

[69] 한편, 1010 동작에서 1차 모델과 2차 모델의 상관도가 임계값을 초과하는

것으로 판단되면, 전자 장치(100)는 1020 동작에서 2차 모델을 이식 모델로 결정할 수 있다. 즉 1차 모델과 2차 모델이 일정 수준 이상으로 유사하면, 프로세서(140)는 2차 모델을 이식 모델로 결정할 수 있다. 이를 통해, 이식 모델이 인간의 지능에 대한 인공 지능으로서 이식될 수 있다. 이 때 이식 모델이 전자 기기, 예컨대 로봇에 이식됨에 따라, 이식 모델에 따른 인공 지능이 인간과 유사하게 작업을 수행하거나, 문제를 해결할 수 있다.

[70]

[71] 다양한 실시예들에 따르면, 인간의 지능과 유사한 인공 지능이 개발될 수 있다. 인간의 지능에 대한 고위수준 지표인 정밀 행동프로파일을 모사할 수 있는 이식 모델이 개발되고, 과적합의 위험 없이, 이식 모델이 인공 지능으로 이식될 수 있다. 이에 따라, 인공 지능이 인간의 행동 프로파일을 복원 가능하여, IoT 분야를 포함하는 인공 지능 비서와 같은 인간 보조 시스템 전반에 인간의 행동 범주 내에서 인간의 행동을 이해 및 예측할 수 있다.

[72]

[73] 다양한 실시예들은, 후술되는 다양한 분야들에 적용 및 응용될 수 있다.

[74]

1. 인간-로봇/컴퓨터 상호작용 분야: 인간의 작업수행/문제해결에 동반되는 행동은 고차원적인 인지 과정에 근거하여 일어나므로, 인간의 행동을 예측하여 활용할 가치가 있는 모든 분야에서 응용될 수 있다. 예로, 감정 컴퓨팅 (affective computing) 분야에서는 인간의 인지 상태의 종류 중 하나인 감정을 읽어 내어 상황에 맞게 인간의 행동을 보조하는 것을 목적으로 한다. 본 시스템은 단순히 감정을 읽어내는 것을 넘어서 컴퓨터가 인식할 수 있는 감정과 맥락적으로 유사한 다른 인지 상태 (예: 각성과 비각성)의 예측을 통해서 인간 행동의 보조에 있어서 효율적으로 대응하는 시스템을 구축하여 인간이 훌륭한 성과를 거둘 수 있도록 보조할 수 있다. 또한 이 기술은 인간-로봇/컴퓨터 상호작용을 포함하는 모든 응용에 기반 기술로 사용될 수 있다. 인간의 준최적(suboptimal) 의사결정 과정을 모방하므로 최적(optimal) 인공지능에 비해 인간과 보다 자연스러운 상호작용을 가능케 한다.

[75]

2. 스마트 IoT 분야: 특히, Internet-of-things (IoT) 분야에서는 다양한 기기를 컨트롤 해야 하므로 각 기기의 컨트롤에 활용되는 인지 기능이 다양할 수 있다. 이때 본 시스템의 범용성은 각 기기를 제어함에 있어서 요구되는 인지 상태의 종류 차이에 관계없이 인간을 보조할 수 있을 뿐 만 아니라, 이미 구축된 IoT 생태계에 새로운 기기가 포함이 되었을 때도 과적합 없이 행동을 예측할 수 있는 AI를 개발할 수 있다.

[76]

3. 전문가 프로파일링 및 스마트 교육 분야: 핵심 고위 인지 과정은 인간의 작업 수행 지능과도 직결되므로, 본 기술을 통해 복잡한 의사결정이 중요한 판사, 의사, 금융 전문가, 군사 작전 지휘관 등에 대한 작업 수행능력 프로파일링이 가능하다. 또한 스마트 교육을 위한 맞춤형 시스템을 위한 사전 프로파일링이 가능하다. 더 나아가 작업 수행 능력 모니터링을 통한 작업 수행 능력 향상도

가능하다.

- [77] 4. AI-인간 공진화형 Application 분야: 인간 지능에 대한 이해는 인간의 신경 수준에서 보상을 최대화하기 위한 의사결정 과정을 이해하는 것에도 적용된다. 기존의 AI는 이러한 인간의 의사결정 과정에 대한 이해가 존재하지 않으나, 인간의 행동 특성을 그대로 예측하는 AI의 개발을 통해 로봇틱스 분야에서는 인간의 행동을 더 잘 예측하는 AI를 개발할 수 있으며, 게임 분야에서는 더욱 지능적인 AI 엔진을 개발할 수 있다.
- [78] 5. 유저 타겟형 AD 분야: 현재 광고 자동 추천 기술은 인간의 과거 검색 기록을 바탕으로 새로운 광고를 추천하고 있다. 그러나 이러한 광고 제안 기술은 개별 인간의 행동 특성에 대한 이해가 결여되어 있어 사용자의 관심범위와 완전히 동떨어진 광고를 제안하는 경우가 많다. 본 기술을 활용하면 사용자의 행동/인지에 보다 직접적인 영향을 끼치는 광고를 추천할 수 있으므로 광고의 효율성을 극대화 시킬 수 있다.
- [79]
- [80] 다양한 실시예들에 따른 전자 장치(100)의 동작 방법은, 태스크에 대한 인간의 처리 데이터를 기반으로, 1차 모델을 피팅하는 동작, 태스크에 대한 1차 모델의 처리 데이터를 기반으로, 2차 모델을 피팅하는 동작, 및 1차 모델과 2차 모델에 대한 프로파일링을 통해, 2차 모델을 인간의 지능에 대한 이식 모델로 결정하는 동작을 포함할 수 있다.
- [81] 다양한 실시예들에 따르면, 인간의 처리 데이터는, 인간이 태스크를 처리하는 동안 발생하는 행위 데이터 또는 뇌 신호 중 적어도 어느 하나를 포함할 수 있다.
- [82] 다양한 실시예들에 따르면, 이식 모델로 결정하는 동작은, 1차 모델과 2차 모델의 상관도를 검출하는 동작, 및 상관도를 기반으로, 2차 모델을 이식 모델로 결정할지의 여부를 판단하는 동작을 포함할 수 있다.
- [83] 다양한 실시예들에 따르면, 전자 장치(100)의 동작 방법은, 이론적으로 적어도 하나의 환경 변수를 설계하는 동작을 더 포함할 수 있다.
- [84] 다양한 실시예들에 따르면, 1차 모델을 피팅하는 동작은, 환경 변수를 기반으로, 인간의 처리 데이터로부터 1차 모델을 피팅할 수 있다.
- [85] 다양한 실시예들에 따르면, 2차 모델을 피팅하는 동작은, 환경 변수를 기반으로, 1차 모델의 처리 데이터로부터 2차 모델을 피팅할 수 있다.
- [86] 다양한 실시예들에 따르면, 1차 모델을 피팅하는 동작은, 인간의 처리 데이터를 기반으로, 1차 모델을 학습하는 동작을 포함하고, 이로 인해, 1차 모델의 행동 프로파일 또는 적어도 하나의 파라미터 중 적어도 어느 하나가 환경 변수를 기반으로 검출될 수 있다.
- [87] 다양한 실시예들에 따르면, 2차 모델을 피팅하는 동작은, 1차 모델의 처리 데이터를 기반으로, 2차 모델을 학습하는 동작을 포함하고, 이로 인해, 2차 모델의 행동 프로파일 또는 적어도 하나의 파라미터 중 적어도 어느 하나가 환경 변수를 기반으로 검출될 수 있다.

- [88] 다양한 실시예들에 따르면, 상관도를 검출하는 동작은, 1차 모델의 행동 프로파일과 2차 모델의 행동 프로파일을 비교하여, 프로파일 상관도를 검출하는 동작, 또는 1차 모델의 파라미터와 2차 모델의 파라미터를 비교하여, 파라미터 상관도를 검출하는 동작 중 적어도 어느 하나, 및 프로파일 상관도 또는 파라미터 상관도 중 적어도 어느 하나를 기반으로, 상관도를 검출하는 동작을 포함할 수 있다.
- [89] 다양한 실시예들에 따르면, 이식 모델로 결정할 지의 여부를 판단하는 동작은, 상관도가 미리 설정되는 임계값을 초과하면, 2차 모델을 이식 모델로 결정하는 동작을 포함할 수 있다.
- [90] 다양한 실시예들에 따르면, 환경 변수는, 상태-전이 불확실성, 상태-공간 복잡성, 신규성, 상태 예측 오류 또는 보상 예측 오류 중 적어도 어느 하나를 포함할 수 있다.
- [91]
- [92] 다양한 실시예들에 따른 전자 장치는, 메모리(130), 및 메모리(130)와 연결되고, 메모리(130)에 저장된 적어도 하나의 명령을 실행하도록 구성된 프로세서(140)를 포함할 수 있다.
- [93] 다양한 실시예들에 따르면, 프로세서(140)는, 태스크에 대한 인간의 처리 데이터를 기반으로, 1차 모델을 피팅하고, 태스크에 대한 1차 모델의 처리 데이터를 기반으로, 2차 모델을 피팅하고, 1차 모델과 2차 모델에 대한 프로파일링을 통해, 2차 모델을 인간의 지능에 대한 이식 모델로 결정하도록 구성될 수 있다.
- [94] 다양한 실시예들에 따르면, 인간의 처리 데이터는, 인간이 태스크를 처리하는 동안 발생하는 행위 데이터 또는 뇌 신호 중 적어도 어느 하나를 포함할 수 있다.
- [95] 다양한 실시예들에 따르면, 프로세서(140)는, 1차 모델과 2차 모델의 상관도를 검출하고, 상관도를 기반으로, 2차 모델을 이식 모델로 결정할 지의 여부를 판단하도록 구성될 수 있다.
- [96] 다양한 실시예들에 따르면, 프로세서(140)는, 이론적으로 적어도 하나의 환경 변수를 설계하고, 환경 변수를 기반으로, 인간의 처리 데이터로부터 1차 모델을 피팅하고, 환경 변수를 기반으로, 1차 모델의 처리 데이터로부터 2차 모델을 피팅하도록 구성될 수 있다.
- [97] 다양한 실시예들에 따르면, 프로세서(140)는, 인간의 처리 데이터를 기반으로, 1차 모델을 학습하도록 구성되고, 이로 인해, 1차 모델의 행동 프로파일 또는 적어도 하나의 파라미터 중 적어도 어느 하나가 환경 변수를 기반으로 검출될 수 있다.
- [98] 다양한 실시예들에 따르면, 프로세서(140)는, 1차 모델의 처리 데이터를 기반으로, 2차 모델을 학습하도록 구성되고, 이로 인해, 2차 모델의 행동 프로파일 또는 적어도 하나의 파라미터 중 적어도 어느 하나가 환경 변수를 기반으로 검출될 수 있다.

- [99] 다양한 실시예들에 따르면, 프로세서(140)는, 1차 모델의 행동 프로파일과 2차 모델의 행동 프로파일을 비교하여, 프로파일 상관도를 검출하고, 1차 모델의 파라미터와 2차 모델의 파라미터를 비교하여, 파라미터 상관도를 검출하고, 프로파일 상관도 또는 파라미터 상관도 중 적어도 어느 하나를 기반으로, 상관도를 검출하도록 구성될 수 있다.
- [100] 다양한 실시예들에 따르면, 프로세서(140)는, 상관도가 미리 설정되는 임계값을 초과하면, 2차 모델을 이식 모델로 결정하도록 구성될 수 있다.
- [101] 다양한 실시예들에 따르면, 환경 변수는, 상태-전이 불확실성, 상태-공간 복잡성, 신규성, 상태 예측 오류 또는 보상 예측 오류 중 적어도 어느 하나를 포함할 수 있다.
- [102]
- [103] 강화학습(Reinforcement Learning, RL)의 급속한 발전은 다양한 유형의 복잡한 문제를 해결하기 위한 알고리즘 개발에 큰 잠재력을 제공했다. 예를 들어, 계층 구조는 희소 보상으로 효과적인 탐구를 촉진하는 것으로 입증되었다. 모델 기반 RL(model-based RL)은 많은 상황에서 샘플 효율을 개선할 수 있는 능력을 입증했다. RL 알고리즘도 생물학적 관련성을 확립해 인간다운 지능을 가진 모델 구축에 대한 낙관론을 키웠다. 다양한 과제를 해결할 수 있는 역량에도 불구하고 샘플 효율성 향상, 적응성, 일반화 등 몇 가지 핵심 과제가 남아 있다. 예를 들어, RL 알고리즘은 환경의 구조를 빠르게 학습할 수 있는 능력이 부족하다. 게다가, 그 행동 정책은 종종 매우 편향적이어서 변화하는 환경에 적응하거나 그것의 작업 지식을 일반 상황에 전달하기 어렵다.
- [104] 이전의 연구에서는 가치 기반 의사결정이 보상 예측 오류(Reward Prediction Error, RPE)에 의해 유도되며, 중간 뇌 도파민 뉴런은 이러한 정보를 암호화한다는 것을 보여주었다. 후자의 연구는 인간의 뇌가 actor-critic 방식을 실행하는 것처럼 보인다는 것을 발견했다. 이러한 연구는 뇌가 경험에서 배우는 방식이 모델 없는 RL(model-free RL)과 유사하다는 생각을 뒷받침한다. 말하자면, 단일 모델 없는 RL은 행동과 신경 데이터의 비교적 작은 변동성을 설명할 수 있다. 이 관습적인 견해는 뇌가 하나 이상의 RL을 구현한다는 생각에 의해 도전받았다. 실제로 인간의 뇌는 모델 없는 RL과 모델 기반 RL을 결합할 수 있을 뿐만 아니라, 문맥 변화에 따라 다른 전략보다 한 전략을 적응적으로 선택할 수 있다. 이러한 적응 과정은 측면 전두엽 피질의 일부에 의해 유도되는 것으로 확인되었으며, 이는 모델 없는 RL 및 모델 기반 RL 전략에 의해 각 예측의 신뢰성을 중합한다. 또한 뇌는 모델 없는 RL과 같이 계산적으로 덜 비싼 전략을 추구하는 경향이 있는데, 특히 매우 불안정하거나 휘발성이 높은 환경에서는 더욱 그러하다. 반면, 전두엽 피질은 성능 신뢰성을 떨어뜨려 모델 기반 학습의 샘플 효율을 획기적으로 향상시키는 데 관여한다. 이는 뇌가 성능, 샘플 효율성 및 계산 비용 사이의 절충을 처리할 수 있는 선천적인 능력을 가지고 있음을 의미한다. 비판적으로, 그것은 두뇌가 환경의 새로운 도전에 가장

잘 대응하는 방법으로 학습 전략을 탐구한다는 이론적 암시로 이어진다.

- [105] 적응 RL에 대한 두뇌와 알고리즘 솔루션 사이에는 몇 가지 공통점이 있지만, 실질적인 차이는 여전히 그들이 문제에 접근하는 방식에 있다. 더욱이 RL의 난제를 효과적으로 다룰 수 있는 뇌의 능력은 RL 알고리즘에 의해 완전히 개발되지 않았다. 이로 인해 다음과 같은 흥미로운 질문이 제기된다. RL 모델이 인간의 행동 데이터에서 인간 RL에 대한 정보를 직접 수집할 수 있는가? 그렇다면 이 모사 모델들은 인간과 유사한 정책을 가지고 있을까? 많은 작품들이 모사를 통해 정책 학습의 효과를 성공적으로 입증했지만, 그들의 정책이 인간의 잠재 정책과 유사한지, 혹은 정책이 다른 과제에 일반화될 수 있는지에 대해서는 거의 알려져 있지 않다. 또 다른 잠재적인 이슈는 과적합이다. 특히, 인간 행동의 회복성을 조사하는 최근의 연구에서는 모델이 원래 적합했던 인간 행동 데이터를 바탕으로 연구 결과를 복제하지 못하는 경우가 종종 있다는 것을 보여주었다. 이는 컴퓨터 모델의 학습된 행동 정책이 인간 RL의 선천적 에너지를 완전히 반영하지 못한다는 것을 시사한다.

[106]

- [107] 현재 강화학습(RL) 알고리즘은 일부 문제에 대해서 인간 지능을 뛰어넘는 해결 능력을 보이지만, 아래와 같은 측면에서는 인간의 강화학습이 우수하다.

- [108] 인간의 강화학습은 데이터 수가 부족하여도 비교적 잘 학습되는 최소 지도(minimal supervision) 학습이 가능하며, 생물의 인지능력(cognitive resource) 한계에 대응하여 낮은 에너지 소모와 높은 성능을 보이는 고효율 학습이 일어난다. 이러한 학습 능력으로 인해 인간의 강화학습은 궁극적으로 다양한 작업(multi-task)으로의 일반화(generalization)를 가능하게 한다.

- [109] 아래의 다양한 실시예들은 자율적, 고효율, 일반화 능력을 갖는 인간 모사형 강화학습 알고리즘의 설계에 필수적인 다각적 정량화 프로세스를 제안한다.

- [110] - 프로세스 1. 정책 신뢰도 정량화 프로세스: 문맥 의존적인 인간의 강화학습 행동 데이터는 매우 복잡한 시간-공간적 상관관계를 가지고 있어 역 강화학습 과정에서 과적합이 일어나기 쉽다. 이를 방지하기 위해 다음과 같이 강화학습 알고리즘의 정책 신뢰도를 정량화한다. 작업 매개변수와 인간의 행동 프로파일간의 매핑 함수를 근사화하고, 작업 매개변수와 강화학습 알고리즘의 행동 프로파일간의 매핑 함수를 근사화한 후, 두 매핑 함수를 비교하는 정량화 프로세스(도 14의 (b))를 실행한다.

- [111] - 프로세스 2. 일반화 능력 검증 프로세스: 인간의 강화학습 과정 모사형 알고리즘의 궁극적인 목적인 일반화 능력의 정밀한 검증을 위해, 실제 문제의 복잡도와 문맥 변화를 매개변수화 시킨 연속적 작업공간에서 샘플링된 일련의 작업에 대한 성능(작업 일반화 가능성)을 검증하는 프로세스(도 14의 (c))를 제공한다.

- [112] - 프로세스 3. 문제해결 정보처리 효율 정량화 프로세스: 문맥 변화에 따라 문제해결 정책을 변화시키는 인간 강화학습 모사형 알고리즘의 적응

능력(앞에서 설명한 정책 신뢰도 정량화 프로세스(프로세스 1)로 정량화)과 다양한 종류의 문제해결을 위한 일반화 능력(앞에서 설명한 일반화 능력 검증 프로세스(프로세스 2))와의 “유기적 연결성”을 확인하기 위하여 마르코프 체인(Markov chain) 관점에서의 정량화(에피소드 인코딩 효율) 프로세스를 제공한다. 문제해결 과정에서 생기는 과거 에피소드가 강화학습 정책에 반영되는 정보 압축 효율과 강화학습 정책으로부터 도출되는 행동의 최적성 비율을 에피소드-정책-행동으로 이어지는 마르코프 체인상의 상호 정보량(mutual information)을 이용해 계산한다. 이 비율은 최적의 문제해결/작업수행을 위해 에피소드 정보를 강화학습 정책결정에 반영하는 정보전달 체계를 나타내는 지표가 된다.

- [113] 상기 3가지 프로세스는 모두 기존에 없는 새로운 기술이다. 본 발명은 “일반화 가능한 인간의 강화학습 능력”을 과적합 없이 알고리즘화 할 수 있음을 실제로 보인 최초의 사례이다.
- [114] 이러한 일련의 프로세스를 통해 과적합 없는 고신뢰도의 일반화 가능한 인간의 강화학습 모사형 알고리즘 설계가 가능함을 보였으며, 또한 이는 기존의 단순한 역 강화학습 과정만으로는 구현할 수 없음을 보였다.
- [115] 프로세스 1의 지표인 정책 신뢰도(reliability) 측면에서는, 최신 강화학습 알고리즘 대비 5배 이상 향상시킬 수 있다. 프로세스 2의 지표인 일반화 능력(generalizability)을 12.8% 향상시킬 수 있다. 프로세스 3의 지표인 에피소드 인코딩 효율 대비 최적행동 효과를 약 100% 향상시킬 수 있다. 이는 아래에서 제안 기술을 이용한 실증연구 결과를 통해 보다 상세히 설명한다.
- [116]
- [117] 강화학습 알고리즘은 생물의 도파민 시스템과 유사하게 가치 기반(value-based)의 학습을 통해 문제를 해결한다. 최근의 연구에서는 딥러닝 기반의 강화학습 알고리즘(에컨대 알파고, 알파제로 등)이 등장하여 바둑과 같이 복잡한 문제에 대해서도 인간의 지능을 뛰어넘는 성능을 보여준다. 그러나 이러한 고성능 강화학습 알고리즘은 인간 지능의 특성을 전부 놓치고 있기에 그 성능에 한계가 명확히 존재한다.
- [118] 일반적인 인공지능 강화학습 알고리즘은 학습에 있어서 많은 데이터를 필요로 하고, 효율보다는 성능을 높이는 것을 목표로 하며, 특정 문제 상황을 해결하는 것에 특화되어 있어 다양한 문제로의 일반화가 불가능하다. 반면, 인간의 강화학습 과정은 이와 반대로 적은 데이터 수에 비해 학습 가능한 뛰어난 최소 지도 학습(minimal supervision learning)의 특성이 있으며, 생물학적 인지능력의 한계에 따라 에너지 소비를 줄이며 학습하는 고효율의 특성이 있고, 특히 특정 문제 상황에만 국한되지 않고 다양한 상황에 대한 일반적 지능을 갖는 특성이 있다.
- [119] 이와 같은 인간의 강화학습 과정의 장점만을 인공지능 강화학습 알고리즘으로 이식하기 위해서는 다음과 같은 접근 방법이 필요하다. (1) 인간 강화학습 모사형

강화학습 알고리즘을 최적화한다. (2) 강화학습 알고리즘의 인간 지능적 특징을 확인(행동 수준)한다: 해당 강화학습 알고리즘을 통해 시뮬레이션된 행동은 인간 지능의 행동과 유사한 형태를 보이는지 다양한 행동 프로파일을 통해 직접적으로 비교할 수 있다. (3) 강화학습 알고리즘의 인간 지능 특징을 확인(매개변수 수준)한다: 각 강화학습 알고리즘을 통해 추출된 시뮬레이션 행동은 다시 각 강화학습 알고리즘으로 재학습되어 매개변수 수준에서의 변화 유무를 통해 인간 지능의 특징을 유지하는지 검증할 수 있다. (4) 정보 이론 수준에서 인간 지능의 특성을 검증한다: 행동과 환경 간의 상호 정보량(mutual information)의 비교를 통해 자연 지능의 특성을 분석할 수 있다. 특히 상호 정보량은 그 분포를 통해 특정 강화학습 알고리즘이 각 자연 지능의 특성에 대해 얼마나 높은 신뢰도를 갖는 알고리즘인지 분석할 수 있다.

[120] 이와 같이 제안된 본 발명은 인공지능 강화학습 알고리즘에 결여된 인간 지능의 장점을 함양하도록 강화학습 알고리즘을 개발하고 검증하는 기술을 다룬다. 이러한 개발 및 다른 강화학습 알고리즘과의 비교를 통한 검증 방법은 기존에 유사한 연구 사례가 없는 독자적 기술이다.

[121]

[122] 본 발명은 인간의 강화학습 과정이 가진 일반화 능력을 강화학습 알고리즘으로 이식하는데 필수적인 정량화 프로세스를 포함한다. (1) 역 강화학습을 통해 도출된 모델이 작업의 문맥 변화를 정책에 얼마나 반영하고 있는지에 대한 정량화, (2) 매개변수화된 작업공간으로부터 샘플링된 작업들에 대한 일반화 능력 정량화, 마지막으로, (3) 정보 이론 관점에서 환경으로부터 행동으로 연결되는 정보의 전환 및 이동 과정이 핵심적인 인간 지능의 행동 원리를 제대로 반영하고 있는지를 정량화 함으로써, 고 신뢰도의 일반화 가능한 강화학습 알고리즘을 설계할 수 있다.

[123] 도 12는 다양한 실시예들에 따른 일반화 가능한 인간 모사형 강화학습 모델 설계를 위한 정량화 방법을 나타내는 흐름도이다.

[124] 도 12를 참조하면, 다양한 실시예들에 따른 컴퓨터를 통해 수행되는 일반화 가능한 인간 모사형 강화학습 모델 설계를 위한 정량화 방법은, 인간의 강화학습 과정이 가진 일반화 능력을 강화학습 모델로 이식하기 위해, 역 강화학습을 통해 도출된 강화학습 모델이 작업의 문맥 변화를 정책에 얼마나 반영하고 있는지에 대한 정량화를 수행하는 정책 신뢰도 정량화 단계(1210)를 포함할 수 있다.

[125] 또한, 일반화 능력의 정밀한 검증을 위해, 작업의 실제 문제의 복잡도와 문맥 변화를 매개변수화 한 작업공간에서 샘플링된 작업에 대한 일반화 가능성을 검증하는 일반화 능력 검증 단계(1220)를 더 포함할 수 있다.

[126] 또한, 환경으로부터 행동으로 연결되는 정보의 전환 또는 이동 과정이 핵심적인 인간 지능의 행동 원리를 제대로 반영하고 있는지를 정량화하는 문제해결 정보처리 효율 정량화 단계(1230)를 더 포함할 수 있다.

[127] 아래에서 다양한 실시예들에 따른 컴퓨터를 통해 수행되는 일반화 가능한 인간

모사형 강화학습 모델 설계를 위한 정량화 방법의 각 단계를 보다 상세히 설명한다.

[128]

[129] 다양한 실시예들에 따른 일반화 가능한 인간 모사형 강화학습 모델 설계를 위한 정량화 방법은 일반화 가능한 인간 모사형 강화학습 모델 설계를 위한 정량화 장치를 예를 들어 설명할 수 있다.

[130] 도 13은 다양한 실시예들에 따른 일반화 가능한 인간 모사형 강화학습 모델 설계를 위한 정량화 장치를 개략적으로 나타내는 블록도이다.

[131] 도 13을 참조하면, 다양한 실시예들에 따른 일반화 가능한 인간 모사형 강화학습 모델 설계를 위한 정량화 장치(1300)는 정책 신뢰도 정량화부(1310)를 포함하여 이루어질 수 있고, 실시예에 따라 일반화 능력 검증부(1320) 및 문제해결 정보처리 효율 정량화부(1330)를 더 포함할 수 있다.

[132] 정책 신뢰도 정량화 단계(1210)에서, 정책 신뢰도 정량화부(1310)는 인간의 강화학습 과정이 가진 일반화 능력을 강화학습 모델로 이식하기 위해, 역 강화학습을 통해 도출된 강화학습 모델이 작업의 문맥 변화를 정책에 얼마나 반영하고 있는지에 대한 정량화를 수행할 수 있다.

[133] 작업, 즉 인간이 학습을 경험하는 모든 상황에서 인간의 강화학습은 다양한 문맥(예컨대, 환경의 불확실성, 복잡도, 보상 조건 등) 변화에 따라서 특정한 행동 양식을 보이는 식의 정책의 변화를 통해 대응한다. 예를 들어, 환경의 불확실성이 높아지는 문맥 변화가 생기는 경우 인간이 목표 지향적 행동을 보이는 것의 효용성이 없기에 이를 지양하는 정책을 선택한다. 역 강화학습을 통해 인간을 모사한 강화학습 모델 역시 동일한 정책을 보이는지 그것을 검증하는 것이 필요하다. 문맥 변화에 따른 행동 양식 변화(즉, 정책의 변화)를 정량화하기 위한 방법으로는 다양한 방법이 제시될 수 있으나, 대표적으로 회귀 분석을 통해 특정 문맥 변화가 정책 변화에 기여하는 영향을 회귀 계수를 통해 정량화할 수 있다.

[134] 보다 구체적으로, 정책 신뢰도 정량화 단계(1210)는 작업의 작업 매개변수와 인간의 행동 프로파일간의 매핑 함수를 근사화하는 단계, 작업 매개변수와 강화학습 알고리즘의 행동 프로파일간의 매핑 함수를 근사화하는 단계, 및 근사화된 두 개의 매핑 함수를 비교하는 단계를 포함하여 이루어질 수 있다.

[135] 여기서, 강화학습 모델은 인간이 학습한 정책 정보를 신뢰성 있게 인코딩하는 모델 기반 제어와 모델 없는(model-free) 제어를 결합한 계산 모델일 수 있다. 또한, 강화학습 모델은 목표 매칭(goal matching, GM), 행동 복제(behavior cloning, BC) 및 정책 매칭(policy matching, PM)의 학습 방법을 통해 구축될 수 있다. 이는 아래에서 보다 상세히 설명한다.

[136] 일반화 능력 검증 단계(1220)에서, 일반화 능력 검증부(1320)는 일반화 능력의 정밀한 검증을 위해, 작업의 실제 문제의 복잡도와 문맥 변화를 매개변수화 한 작업공간에서 샘플링된 작업에 대한 일반화 가능성을 검증할 수 있다.

- [137] 일반화 능력은 인간이 갖는 학습 특성으로, 한 작업에서 보이는 문맥 변화에 따른 정책 변화 특성을 다른 작업에 있어서도 동일하게 보이는 것이다. 특정한 작업을 학습하고 보상을 최대화하기 위해 보인 인간의 강화학습 특성, 즉 문맥 변화에 따른 정책 변화를 성공적으로 반영한 모델(즉, 단계(1210)을 통해 검증된)은 문제의 복잡도 등 다른 문맥이 변화하는 작업에서도 인간이 보였던 특성을 통해 일반화 가능한 성능을 보이는 것을 확인할 수 있다. 이를 폭 넓게 검증하기 위해 문제의 복잡도 및 문맥 변화를 매개변수화 및 이를 조절하여 다양한 작업을 만들고 이에 노출시켜 그 성능을 통해 일반화 능력을 검증할 수 있다.
- [138] 문제해결 정보처리 효율 정량화 단계(1230)에서, 문제해결 정보처리 효율 정량화부(1330)는 환경으로부터 행동으로 연결되는 정보의 전환 또는 이동 과정이 핵심적인 인간 지능의 행동 원리를 제대로 반영하고 있는지를 정량화할 수 있다.
- [139] 인간 지능의 행동 원리는 자원의 효율적 분배에 있다. 문맥의 변화에 따라, 인지적 노력이 많이 필요하지만 확실한 고성능을 보일 수 있는 목표 지향적 행동을 보일 때도 있고 효율성을 강조한 습관적 행동을 보일 때도 있다. 일반적으로, 인간은 두 정책의 적절한 분배를 통해 고성능이며 고효율인 행동 양식을 갖는다. 이 적절한 정책의 변화가 일어나는 지 정량화하기 위해, 두 종류의 상호정보량(mutual information)을 활용할 수 있다. 첫째는 이전의 경험과 현재의 선택 사이의 상호정보량으로, 이 값이 낮다면 정보의 압축을 통한 효율적 선택으로 이해할 수 있다(효율성 지표). 둘째는 현재의 선택과 현재의 선택지 중 최고의 보상 값을 갖는 선택(최적 선택) 사이의 상호정보량으로, 이 값이 높다면 고성능으로 볼 수 있다(성능 지표). 두 상호정보량의 비율(성능 지표/ 효율성 지표)을 통해 인간 지능의 행동 원리를 복원하는 지 그 정보처리의 효율을 정량화할 수 있다.
- [140] 문제해결 정보처리 효율 정량화부(1330)는 문맥 변화에 따라 문제해결 정책을 변화시키는 인간 모사형 강화학습 모델의 정책 신뢰도 정량화부(1310)를 통한 적응 능력과 일반화 능력 검증부(1320)를 통한 문제해결을 위해 검증된 일반화 능력과의 연결성 확인을 위해 마르코프 체인(Markov chain)을 이용하여 정량화할 수 있다.
- [141] 또한, 문제해결 정보처리 효율 정량화부(1330)는 문제해결 과정에서 발생하는 과거 에피소드가 강화학습 정책에 반영되는 정보 압축 효율과 강화학습 정책으로부터 도출되는 행동의 최적성 비율을 에피소드-정책-행동으로 이어지는 마르코프 체인상의 상호 정보량(mutual information)을 이용하여 계산할 수 있다.
- [142] 여기서, 행동의 최적성 비율은 최적의 문제해결을 위해 에피소드 정보를 강화학습 정책결정에 반영하는 정보전달 체계를 나타내는 지표가 될 수 있다.
- [143] 아래에서 다양한 실시예들에 따른 일반화 가능한 인간 모사형 강화학습 모델

설계를 위한 정량화 방법 및 장치에 대해 보다 상세히 설명한다.

[144]

[145] 심층 강화학습(deep RL) 모델은 최소한의 지도(supervision)로 다양한 유형의 작업을 해결할 수 있는 큰 잠재력을 보여주었지만, 제한된 경험에서 빠르게 학습하고, 환경 변화에 적응하며, 단일 작업에서 학습을 일반화한다는 측면에서 몇 가지 핵심 과제가 남아 있다. 의사결정 신경과학의 최근 증거는 인간의 뇌가 이러한 문제들을 해결할 수 있는 선천적인 능력을 가지고 있다는 것을 보여주었고, 이는 샘플 효율적이고 적응적이며 일반화될 수 있는 RL 알고리즘에 대한 신경과학에서 영감을 받은 해결책 개발에 대한 낙관론으로 이어졌다.

[146]

여기에서는 전두엽(prefrontal) RL이라고 부르는 모델 기반 제어와 모델 없는(model-free) 제어를 적응적으로 결합한 계산 모델이 인간이 학습한 높은 수준의 정책 정보를 신뢰성 있게 인코딩하는 것을 보여주며, 이 모델은 학습된 정책을 광범위한 작업에 일반화할 수 있다.

[147]

먼저, 피험자들이 2단계 마르코프 의사결정 과제를 수행하는 동안 수집된 82명의 피실험자의 데이터에 대해 전두엽 RL, 심층 RL, 메타 RL 알고리즘을 훈련시켰는데, 이 과정에서 목표, 상태-변환 불확실성, 상태-공간 복잡성을 실험적으로 조작했다. 잠재적 행동 프로파일과 매개변수 회복성 시험을 조합한 신뢰도 시험에서, 전두엽 RL이 인간 피험자의 잠재된 정책을 신뢰성 있게 학습한 반면, 다른 모든 모델은 이 시험을 통과하지 못했다는 것을 보여주었다. 둘째, 이러한 모델들이 본래의 작업에서 배운 것을 일반화하는 능력을 실증적으로 시험하기 위해, 그것들을 환경 변동성 문맥에 배치했다. 구체적으로, 10가지 다른 마르코프 의사결정 작업으로 대규모 시뮬레이션을 실행했는데, 이 작업에서 잠재적 문맥 변수는 시간이 지남에 따라 변화한다. 다양한 실시예들에 따른 정보이론적 분석은 전두엽 RL이 가장 높은 수준의 적응성과 성공적 인코딩 효과를 보인다는 것을 알 수 있다. 이것은 두뇌가 일반적인 문제를 해결하는 방법을 모방한 컴퓨터 모델이 기계학습의 주요 난제에 대한 실질적인 해결책으로 이어질 수 있는 가능성을 공식적으로 시험하기 위한 첫 번째 시도이다.

[148]

본 발명은 다음과 같은 근본적인 질문을 검토한다. 알고리즘이 인간으로부터 일반화할 수 있는 정책을 배우는 것이 가능한가? 이를 위해 이 문제를 신뢰도 시험과 경험적 일반화 시험의 전제조건으로 두 가지 공식 시험으로 삼는다. 본 발명의 작업은 다음과 같이 요약된다.

[149]

인간의 잠재 정책 학습. 여기에서는 82명의 피험자의 데이터를 다양한 RL 모델에 장착했는데, 각 모델은 심층 RL, 메타 RL, 전두엽 RL 등 다양한 방식으로 모델 없는 제어 및 모델 기반 제어를 구현한다. 여기서 목표, 상태-변환 불확실성, 상태-공간 복잡성이 실험적으로 조작되는 2단계 마르코프 의사결정 과제를 수행하는 인간 참가자들로부터 수집된 데이터를 사용했다.

[150]

신뢰도 시험. 엄격한 잠재적 행동 프로파일 회복성 시험을 사용하여 모델 기반

제어와 모델 없는 제어(전두엽 RL이라 함)를 적응적으로 결합한 계산 모델의 잠재 정책이 인간 피험자와 질적으로 유사하지만, 다른 모든 모델은 효과를 재현하지 못한다.

[151] 경험적 일반화 능력 시험. 원래 작업에서 학습한 내용을 일반화하는 모델의 능력을 시험하기 위해, 시간에 따라 잠재적 상황 변수가 변화하는 10가지 마르코프 의사결정 작업으로 대규모 시뮬레이션을 실행했다. 여기서 전두엽 RL이 가장 높은 수준의 적응성과 성공적 인코딩 효능을 보인다는 것을 발견했다.

[152] 이 작업은 컴퓨팅 모델이 인간의 잠재된 정책을 신뢰성 있게 학습할 수 있는 가능성을 공식적으로 시험하기 위한 첫 번째 시도이다. 더욱이 이 접근방식은 기계학습의 주요 난제에 대한 실질적인 해결책을 제공하여 보다 인간과 같은 지능을 설계할 수 있게 해준다.

[153]

[154] 인간의 잠재 정책 학습

[155] 도 14는 다양한 실시예들에 따른 인간의 잠재 정책 학습, 신뢰도 시험 및 경험적 일반화 시험을 설명하기 위한 도면이다.

[156] 도 14의 (a)를 참조하면, 인간과 유사한 방식으로 작업을 배우고 수행하는 RL 모델을 구축하기 위해 목표 매칭(goal matching, GM), 행동 복제(behavior cloning, BC), 정책 매칭(policy matching, PM) 등 3가지 훈련 방법을 고려한다. 여기서 인간의 잠재 정책 학습이라고 부르는 이 과정은 인간의 행동 데이터에서 직접 행동 정책을 학습하기 위한 것이다.

[157] RL 모델은 작업 환경과 상호 작용하여 향후 예상되는 보상의 양을 최대화하므로 훈련에는 인간의 행동 데이터를 사용하지 않는다. 그러나 모델을 훈련시키는 데 사용되는 작업(목표)은 인간 피험자가 수행하는 과제와 정확히 같다. 따라서 이 방법을 목표 매칭(GM)이라고 부른다.

[158] 정책 매칭(PM)은 목표 매칭(GM)과 행동 복제(BC)가 결합되어 목표 매칭과 행동 복제를 모두 달성할 수 있다. 구체적으로, RL 모델은 인간이 보상 극대화를 수행하는 방식을 모방하는 방식으로 훈련된다. 각 훈련 에폭(epoch)에는 RL 모델이 보상(목표 매칭)을 극대화하기 위한 작업의 에피소드를 완성하고, 이후 모델의 행동과 인간 대상 행동의 차이를 손실함수(행동 복제)로 환산한다. 이 방법은 이전에 신경 데이터를 설명하기 위한 계산 모델을 훈련하기 위해 사용되었다. 표준 역 RL 방법은 빠른 문맥 변화를 가진 작업에 직접 적용할 수 없기 때문에 여기에서는 표준 역 RL 방법을 고려하지 않음에 주목해야 한다. 실제로 시간이 지남에 따라 보상가치와 환경통계가 모두 변하며, 샘플 크기가 너무 작은 보상함수를 역 RL 방법으로 추정하는 것은 거의 불가능하다(과제당 약 400회의 실험).

[159]

[160] 도 15는 다양한 실시예들에 따른 실험에 사용된 RL 모델의 구조를 설명하기

위한 도면이다.

- [161] 도 15를 참조하면, 실험을 위해 심층 RL, 메타 RL, 전두엽 RL의 세 가지 RL 모델을 사용했다. 첫 번째 유형은 DDQN이라고도 알려진 Double DQN(심층 RL)으로 구현되었다. 그것은 모델 없는 RL에 근접한 대표적인 심층 RL 모델 중 하나이다. 이 모델(각각 GM-DDQN, PM-DDQN)을 훈련하기 위해 목표 매칭과 정책 매칭 방법을 모두 사용했다.
- [162] 두 번째 유형은 메타 RL(meta RL)로 구현되었다. 이 모델은 모델 없는 RL 및 모델 기반 RL을 모두 수용한다. 특히, 메타 RL은 환경 문맥 변화에 적응적으로 반응하는 것으로 알려져 있다. 이 모델(각각 GM-metaRL, PM-metaRL)을 교육하기 위해 목표 매칭과 정책 매칭 방법을 모두 사용했다.
- [163] 세 번째 유형의 RL 모델은 측면 전두엽 피질 및 복측 선조체(전두엽 RL)의 신경 활동을 설명하기 위해 연산 모델로 구현되었다. 이 모델에는 기준 모델과 적응형 모델의 두 가지 버전이 있다. 이 모델들은 모델 없는 RL과 모델 기반 RL 사이에서 동적으로 중재함으로써 작업을 학습한다. 구체적으로는 모델 없는 RL 및 모델 기반 RL 전략에 할당된 제어의 정도를 시험별로 조정하며, 이 하향 조정 신호는 각 RL 전략의 예측 신뢰도에 근거하여 계산한다. 정책 매칭 방법을 사용하여 이 두 모델(PM-pfcRL1과 PM-pfcRL2)을 학습하였다. 이전 연구에서는 이러한 모델을 데이터에 적합시키는 데 이 방법이 효과적이지 않다는 것을 밝혀냈기 때문에 이 경우에는 목표 매칭을 사용하지 않았다.
- [164]
- [165] 뇌에서 영감을 받은 RL 모델의 신뢰도
- [166] 도 14의 (b)에 도시된 바와 같이, RL 모델이 인간 행동과 잠재 정책을 얼마나 신뢰성 있게 모방하는지를 평가하기 위해 신뢰도 시험을 실시했다. 이 시험은 인간이 과제를 수행하면서 학습한 고도의 정책 정보를 인코딩할 수 있는 능력을 검증한다. 이 과정은 잠재 행동 프로파일링과 회복성 시험으로 구성된다.
- [167] 인간이 작업로부터 배우는 잠재 정책을 평가하는 한 가지 일반적인 방법은 잠재적 작업 매개변수(예: 목표 및 상태-변환 불확실성)가 행동에 미치는 영향을 정량화하는 것이다. 이 척도는 학습 에이전트가 환경구조의 변화에 대응해 어떻게 행동을 변화시키는지 반영한다. 각각 주어진 작업 매개변수 θ 와 행동 데이터 x 에 대해 잠재 행동 프로파일 h 는 다음 식과 같이 정의된다.
- [168] [수식1]
- $$x = h(\theta_{Task})$$
- [169] 여기서, h 는 다항식 기능이나 신경망과 같은 매개변수화된 기능일 수 있다. 에이전트의 작업 수행이 문맥 변경과 무관하거나 에이전트가 임의로 선택하는 경우, 효과 크기(즉, h 의 매개변수 값)는 0이 된다. 여기에서는 일반 선형 모델을 h 로 간단하게 사용한다.
- [170] 잠재 동작 프로파일 회복성 시험의 목적은 인간의 잠재 정책과 RL 모델의 잠재

정책 사이의 일관성을 평가하는 것이다. 모델의 매개변수를 인간 피실험자의 데이터 x_{Human} 에 맞춘 후, 원래 작업에서 원래 피팅 모델로 시뮬레이션을 실행하여 시뮬레이션한 데이터 x_{Model} 을 생성한다. 그런 다음 x_{Human} 과 x_{Model} 에 대해 각각 잠재 행동 프로파일링을 실시한다. 이들 두 잠재적 프로파일 간의 유의미한 양의 상관관계는 RL 모델이 학습한 잠재 정책이 인간의 잠재 정책과 유사하다는 것을 나타낸다.

- [171] 신뢰도 시험을 위해, 잠재 행동 프로파일의 회복성을 조사하기 위해 6가지 RL 모델(도 15)과 임의 에이전트를 제어조건으로 하여 일련의 실험을 실시했다. 첫 번째 단계에서는 82명의 피험자 데이터(도 14의 (b)의 x_{Human})에 대한 전두엽 RL, 메타 RL, 심층 RL을 교육했다. 피험자들이 2단계 마르코프 의사결정 작업을 수행하는 동안 데이터 집합이 수집되었다. 두 번째 단계에서는 모든 RL 모델이 동일한 2단계 마르코프 의사결정 작업을 수행하는 다른 시뮬레이션 집합을 실행하여 또 다른 행동 데이터 집합(도 14의 (b)의 x_{Model})을 수집했다. 그런 다음 잠재적 행동 프로파일 h_{Human} , h_{Model} 을 다음 식과 같이 계산했다.

- [172] [수식2]

$$x_{Human} = h_{Human}(\theta_{Task}), x_{Model} = h_{Model}(\theta_{Task})$$

- [173] 여기서, θ_{Task} 는 작업 매개변수를 나타낸다. 이는 1000개 이상의 모델 피팅 공정을 포함한 대규모 실험이다:

- [174] 7(모델) × 82(대상) × 2(훈련 및 재교육)

- [175] 다양한 실시예들에 따른 신뢰도 시험 결과에서, RL 모델과 피험자 대상 간 동작 매칭을 정량화하는 모델 피팅 면에서는 PM-meta RL이 가장 높은 성능을 보였고, 전두엽 RL과 심층 RL이 그 뒤를 이었다. 예상대로 목표 매칭으로 훈련된 RL 모델은 상대적으로 피팅 성능이 떨어지는 것으로 나타났다.

- [176] 그러나 잠재 행동 프로파일의 체계적인 회복 분석에서 전두엽 RL 모델(PM-pfcRL2)의 잠재 행동 프로파일이 인간 피험자와 질적으로 유사한 반면, 다른 모든 RL 모델은 효과를 복제하지 못했다. PM 방법으로 훈련된 메타 RL의 경우 의한 상관관계를 보였지만, 이 모델이 작업을 수행하는 방식이 인간의 그것과 근본적으로 다를 수 있음을 나타내는 음의 상관관계가 있다. 상관관계의 가파름과 유의성을 모두 고려한 적합도 통계량을 계산할 때 이 효과는 더 극적으로 나타난다. 전두엽 RL 모델(PM-pfcRL2)의 효과 크기는 다른 모든 RL 모델의 효과 크기보다 3배 이상 크다. 이러한 결과는 단순히 인간의 행동을 모사하는 것은 에이전트가 실제로 인간의 잠재 정책을 학습하는 것을 의미하지는 않는다는 것을 시사한다.

- [177]

- [178] 뇌에서 영감을 받은 RL 모델의 경험적 일반화 능력

- [179] 도 16은 다양한 실시예들에 따른 각 RL 모델의 일반화 시험을 위한 시뮬레이션 환경을 설명하기 위한 도면이다.

- [180] 도 16을 참조하면, 모델들이 원래 작업에서 다른 작업으로 배운 것을 일반화할 수 있는 능력을 경험적으로 시험하기 위해(도 14의 (c)) 환경 변동성의 문맥에서 모델을 배치했다. 앞에서 설명한 바와 동일한 RL 모델 집합을 사용하여 각각 다른 방식으로 잠재적 상황 변수를 조작하는 10개의 서로 다른 마르코프 의사결정 과제로 대규모 시뮬레이션을 실행했다. 작업은 작업 구조(사다리(Ladder) 및 트리(Tree))와 작업 불확실성(고정(Fixed), 드리프트(Drift), 스위치(Switch), 드리프트 + 스위치(Drift+Switch))의 두 가지 작업 매개변수를 체계적으로 조작해 만들어졌다. 도 16의 (b)에 도시된 바와 같이, 작업 구조는 사다리와 트리 타입을 사용하였다. 도 16의 (c)에 도시된 바와 같이, 작업 불확실성 변동에 대해, 4가지 다른 유형의 상태 전환 함수를 검토했는데, 각각의 상태 전환 확률 값은 시험마다 다른 방식으로 변경되었다.
- [181] 첫 번째 유형("고정")은 고정 상태 변환 확률을 사용한다. 두 번째 유형("드리프트")은 무작위 보행에 따른 상태 변환 확률을 사용하며, 상태 변환 확률 값은 상대적으로 느리게 변경된다. 세 번째 유형("스위치")은 각각 낮은 불확실성과 높은 불확실성을 가진 조건이라는 두 가지 다른 상태 변환 조건 사이에서 번갈아 나타난다. 이 작업에서 학습 에이전트는 작업 구조의 급격한 변화를 경험하며, 신속하게 적응할 필요가 있다. 네 번째 유형("드리프트 + 스위치")은 두 번째 유형과 세 번째 유형의 혼합물이다. 도 16의 (d)에 도시된 바와 같이, 각 작업의 전체 구성을 나타낼 수 있다. Task 1과 Task 10은 뇌의 RL 과정을 조사하는 이전 연구에서 사용된 작업에 해당한다.
- [182] 도 17은 다양한 실시예들에 따른 RL 모델의 적응 능력에 대한 시뮬레이션 결과를 나타내는 도면이다.
- [183] 경험적 일반화 능력을 시험하기 위해, 원래 데이터 집합에 대해 훈련된 6개의 RL 모델(앞에서 언급한 RL 모델)이 10개의 마르코프 의사결정 과제를 수행하는 시뮬레이션을 실행했다. 여기에는 총 4,920개의 시뮬레이션(= 82개 과제(subject) × 6 RL 모델 × 10개 작업(task))이 포함되었다. 모든 작업에 걸친 평균 성능은 경험적 일반화 능력을 나타내며, 각 작업에 대한 성능은 서로 다른 상황에서 해당 모델의 적응 능력을 나타낸다. 도 17을 참조하면, 전두엽 RL 모델이 가장 높은 수준의 일반화 능력을 보인다는 것을 발견할 수 있다.
- [184]

[표 1]

	Task 1	Task 2	Task 3	Task 4	Task 5	Task 6	Task 7	Task 8	Task 9	Task 10	Success rate
PM-DQN	FAIL	FAIL	FAIL	FAIL	FAIL	FAIL	FAIL	FAIL	FAIL	FAIL	0/10
GM-DDQN	FAIL	FAIL	FAIL	FAIL	FAIL	FAIL	FAIL	FAIL	FAIL	FAIL	0/10
PM-metaRL	FAIL	FAIL	0.35	FAIL	FAIL	0.36	0.59	0.59	FAIL	FAIL	4/10
GM-metaRL	FAIL	FAIL	0.38	FAIL	FAIL	0.36	0.55	0.55	0.51	0.52	6/10
PM-pfcRL1	FAIL	FAIL	0.42	FAIL	FAIL	0.36	0.71	0.71	0.60	0.60	6/10
PM-pfcRL2	FAIL	0.51	0.40	0.51	0.52	0.38	0.71	0.71	0.60	0.60	9/10

[185] 특히, 표 1을 참조하면, PM-pfcRL2는 10개 작업 중 9개 작업을 성공적으로 해결하고, 9개 작업 중 8개 작업에서 정규화 보상으로 가장 높은 점수를 받았다. GM-metaRL과 PM-pfcRL1은 두 번째로 좋은 성능을 보였다. PM-pfcRL1의 성능은 GM-metaRL과 동일했지만 PM-pfcRL1은 6개 작업 중 5개 작업에서 월등히 우수한 성적을 거뒀다. 이러한 결과를 종합하면, 전두엽 RL 모델(PM-pfcRL1 및 PM-pfcRL2)이 원래 작업에서 배운 것을 일반화할 수 있는 최고의 능력을 가지고 있음을 시사한다.

[186] RL 모델의 일반화 능력을 정량화하기 위해 잠재적 정보-이론적 척도를 제공할 수 있다. 일반화 능력의 성격을 보다 잘 이해하기 위해 정보이론적 분석을 실시했다. 이 분석은 (1) 사건의 과거 에피소드의 관측에서 RL 모델의 작용으로 전달되는 정보의 양과, (2) 그 작용의 최적성의 정도를 정량화하기 위해 설계되었다. 일반성이 높을수록 RL 모델이 에피소드 정보를 보다 효과적으로 인코딩하여 최적의 작용을 발생시킨다는 가설을 세웠다. 이와 같이, 모델의 일반화 능력을 (1) 에피소드 사건 및 에이전트의 행동("에피소드 인코딩 효과")에서 얻은 상호 정보뿐만 아니라, (2) 에이전트의 행동과 최적의

행동("선택적 최적성")으로 정량화할 수 있을 것으로 기대한다. 최적의 행동은 작업의 매개변수 변경에 대해 충분히 알고 있다고 가정하여 이상적인 에이전트가 취한 행동으로 정의되었다. 에피소드 인코딩 효과는 $I(a_t; a_t^*)$ 로

정의되며, 여기서 F_{t-1} 과 a_t 는 시도 t-1에서의 에피소드 변수 및 시도 t에서의 행동이다. 선택 최적성은 $I(a_t; a_t^*)$ 로 정의되며, 여기서 a_t 와 a_t^* 는 각각 RL

에이전트와 이상적인 에이전트의 선택(행동)이다. 여기서 일반화할 수 있는 RL 에이전트의 한 가지 기본적인 요구사항은 과거 에피소드에서 그것의 행동과 작업 수행으로 정보를 전송하는 능력이라고 가정했다. 따라서 "episodic encoding efficacy"라고 불리는 에피소드 인코딩 효과와 선택적 최적성의 상관관계는 RL 모델의 일반화 능력을 나타내는 하나의 잠재적 정보-이론적 지표가 될 수 있다.

- [187] 다양한 실시예들에 따른 에피소드 인코딩 효과 검증을 위해, 그런 다음, 이 척도를 사용하여 비율 $I(F_{t-1}; a_t)/I(a_t; a_t^*)$ 및 에피소드 인코딩 효과의

대용으로서 적합도 통계량을 계산했다. 여기서 전두엽 RL(PM-pfcRL1과 PM-pfcRL2 둘 다)이 가장 높은 수준의 에피소드 인코딩 효과를 보인다는 것을 발견했다. 특히, 가장 일반화할 수 있는 모델인 PM-pfcRL2는 10개 작업 중 8개 작업에서 에피소드 인코딩 효과와 선택적 최적성 사이에 유의미한 상관관계를 보였다. 또한, 경험적 일반화 능력(도 17)은 대부분 성공적 인코딩 효과의 R2와 일치한다는 점에 주목한다. 이러한 결과는 3가지 중요한 의미를 갖는다. 첫째, 에피소드 인코딩 효과는 일반화 능력의 성격을 더 잘 이해하는데 도움을 준다. 둘째, 에피소드 인코딩 효과는 에이전트의 일반화 능력을 계량화할 수 있는 좋은 후보가 될 수 있다. 이 척도는 매우 일반화할 수 있는 RL 알고리즘 설계에 직접 사용될 수 있다.

[188]

- [189] 다양한 실시예들에 따르면, 인간의 강화학습 과정이 가진 일반화 능력을 강화학습 알고리즘으로 과적합 없이 알고리즘화 할 수 있는, 일반화 가능한 인간 모사형 강화학습 알고리즘 설계를 위한 정책 신뢰도, 정보처리 효율 및 일반화 능력의 정량화 방법 및 장치를 제공할 수 있다.

[190]

- [191] 다양한 실시예들은 인간 지능의 모든 행동은 고차원적인 인지 기능에 근거하여 일어나므로, 이 행동을 예측하여 활용할 가치가 있는 모든 분야에서 응용될 수 있다. 일례로, 인간의 문맥 의존적인 강화학습 과정을 모사하는 모델을 이용하여 인간 행동의 보조에 있어서 효율적으로 대응하는 시스템을 구축하여 인간이 훌륭한 성과를 거둘 수 있도록 보조할 수 있다.

[192]

사물인터넷(Internet-of-Things, IoT) 분야에서는 다양한 기기를 컨트롤 해야 하므로 각 기기의 컨트롤에 활용되는 인지 기능이 다양할 수 있다. 이 때, 다양한 실시예들에 따른 시스템의 범용성은 각 기기를 제어함에 있어서 요구되는 인지

상태의 종류 차이에 관계없이 인간을 보조할 수 있을 뿐만 아니라, 이미 구축된 IoT 생태계에 새로운 기기가 포함이 되었을 때도 과적합 없이 행동을 예측할 수 있는 AI를 개발할 수 있다.

[193] 또한, 다양한 문제로의 일반화 능력은 인간의 작업 수행 지능과도 직결되므로, 다양한 실시예들에 따른 기술을 통해 복잡한 의사결정이 중요한 판사, 의사, 금융 전문가, 군사 작전 지휘관 등에 대한 작업 수행능력 프로파일링이 가능하다. 또한 스마트 교육을 위한 맞춤형 시스템의 기반 기술로도 활용이 가능하다.

[194] 다양한 실시예들에 따른 기술을 이용하여 도출되는 인간의 강화학습 모사형 알고리즘은 인간의 의사결정의 핵심과정을 이해하는 도구로도 활용될 수 있다. 기존의 AI는 이러한 인간의 의사결정 과정에 대한 이해가 존재하지 않으나, 인간의 행동 특성을 그대로 예측하는 AI의 개발을 통해 로보틱스 분야에서는 인간의 행동을 더 잘 예측하고 보조하는 AI를 개발할 수 있으며, 게임 분야에서는 인간과 자연스러운 상호작용이 가능한 더욱 지능적인 AI 엔진을 개발할 수 있다.

[195] 한편, 현행 광고 제안 기술은 인간의 과거 검색 기록을 바탕으로 새로운 광고를 추천하고 있다. 그러나 이러한 광고 제안 기술은 개별 인간의 행동 특성에 대한 이해가 결여되어 있어 사용자의 관심범위와 완전히 동떨어진 광고를 제안하는 경우가 많다. 다양한 실시예들에 따른 기술을 활용하면 인간-AI 간의 공진화를 통해 사용자의 행동 범위 내에 존재하는 광고를 추천할 수 있다.

[196] 이상과 같이, 인간 지능의 특성을 함양하도록 하는 인간 모방형 인공지능의 설계는 단순히 인간의 행동을 더욱 유사하게 예측할 수 있을 뿐만 아니라, 그 특성이 학습과 성능의 효율에 있으므로 더욱 적은 노력으로 더 나은 결과를 얻을 수 있다는 점에서 인공지능 산업 전반에 적용 가능한 유익한 기술이다. 특히, 강화학습은 문제 해결 및 의사결정에 큰 도움이 되므로 인간을 포함한 지능적 판단이 필요한 모든 인공지능 개발에 중요하다.

[197] 인공지능의 개발은 특정 문제 상황에 대한 해결을 위해 상당한 계산과 시간 자원이 투자됨에도 불구하고 그 인공지능이 다양한 문제 해결이 아닌 특정 문제 해결에만 적용 가능하다는 큰 단점이 있다. 이와 반대로 본 시스템은 일반화 가능한 알고리즘의 개발이 가능해 다양한 문제 해결에 적용될 수 있다.

[198] 개발 중 및 개발된 모든 인공지능의 자연 지능적 특성 검증에 적용 가능하다. 인간 지능을 모사하여 인간의 인지 과정을 예측하고자 하는 모델은 쉽게 과적합의 오류에 빠지기 때문에 반드시 이러한 과적합의 오류를 제거해야 한다.

[199]

[200] 다양한 실시예들에 따른 컴퓨터를 통해 수행되는 일반화 가능한 인간 모사형 강화학습 모델 설계를 위한 정량화 방법은, 인간의 강화학습 과정이 가진 일반화 능력을 강화학습 모델로 이식하기 위해, 역 강화학습을 통해 도출된 강화학습 모델이 작업의 문맥 변화를 정책에 얼마나 반영하고 있는지에 대한 정량화를

수행하는 정책 신뢰도 정량화 단계를 포함할 수 있다.

- [201] 다양한 실시예들에 따르면, 정책 신뢰도 정량화 단계는, 작업의 작업 매개변수와 인간의 행동 프로파일간의 매핑 함수를 근사화하는 단계, 작업 매개변수와 강화학습 알고리즘의 행동 프로파일간의 매핑 함수를 근사화하는 단계, 및 근사화된 두 개의 매핑 함수를 비교하는 단계를 포함할 수 있다.
- [202] 다양한 실시예들에 따르면, 상기 정량화 방법은, 일반화 능력의 정밀한 검증을 위해, 작업의 실제 문제의 복잡도와 문맥 변화를 매개변수화 한 작업공간에서 샘플링된 작업에 대한 일반화 가능성을 검증하는 일반화 능력 검증 단계를 더 포함할 수 있다.
- [203] 다양한 실시예들에 따르면, 상기 정량화 방법은, 환경으로부터 행동으로 연결되는 정보의 전환 또는 이동 과정이 핵심적인 인간 지능의 행동 원리를 제대로 반영하고 있는지를 정량화하는 문제해결 정보처리 효율 정량화 단계를 더 포함할 수 있다.
- [204] 다양한 실시예들에 따르면, 문제해결 정보처리 효율 정량화 단계는, 문맥 변화에 따라 문제해결 정책을 변화시키는 인간 모사형 강화학습 모델의 정책 신뢰도 정량화를 통한 적응 능력과 문제해결을 위해 검증된 일반화 능력과의 연결성 확인을 위해 마르코프 체인(Markov chain)을 이용하여 정량화 할 수 있다.
- [205] 다양한 실시예들에 따르면, 문제해결 정보처리 효율 정량화 단계는, 문제해결 과정에서 발생하는 과거 에피소드가 강화학습 정책에 반영되는 정보 압축 효율과 강화학습 정책으로부터 도출되는 행동의 최적성 비율을 에피소드-정책-행동으로 이어지는 마르코프 체인상의 상호 정보량(mutual information)을 이용하여 계산할 수 있다.
- [206] 다양한 실시예들에 따르면, 행동의 최적성 비율은 최적의 문제해결을 위해 에피소드 정보를 강화학습 정책결정에 반영하는 정보전달 체계를 나타내는 지표가 될 수 있다.
- [207] 다양한 실시예들에 따르면, 강화학습 모델은, 인간이 학습한 정책 정보를 신뢰성 있게 인코딩하는 모델 기반 제어와 모델 없는(model-free) 제어를 결합한 계산 모델일 수 있다.
- [208] 다양한 실시예들에 따르면, 강화학습 모델은, 목표 매칭(goal matching, GM), 행동 복제(behavior cloning, BC) 및 정책 매칭(policy matching, PM)의 학습 방법을 통해 구축될 수 있다.
- [209]
- [210] 다양한 실시예들에 따른 일반화 가능한 인간 모사형 강화학습 모델 설계를 위한 정량화 장치(1300)는, 인간의 강화학습 과정이 가진 일반화 능력을 강화학습 모델로 이식하기 위해, 역 강화학습을 통해 도출된 강화학습 모델이 작업의 문맥 변화를 정책에 얼마나 반영하고 있는지에 대한 정량화를 수행하는 정책 신뢰도 정량화부(1310)를 포함할 수 있다.
- [211] 다양한 실시예들에 따르면, 정책 신뢰도 정량화부(1310)는, 작업의 작업

매개변수와 인간의 행동 프로파일간의 매핑 함수를 근사화하고, 작업 매개변수와 강화학습 알고리즘의 행동 프로파일간의 매핑 함수를 근사화한 후, 근사화된 두 개의 매핑 함수를 비교할 수 있다.

- [212] 다양한 실시예들에 따르면, 상기 정량화 장치(1300)는, 일반화 능력의 정밀한 검증에 의해, 작업의 실제 문제의 복잡도와 문맥 변화를 매개변수화 한 작업공간에서 샘플링된 작업에 대한 일반화 가능성을 검증하는 일반화 능력 검증부(1320)를 더 포함할 수 있다.
- [213] 다양한 실시예들에 따르면, 상기 정량화 장치(1300)는, 환경으로부터 행동으로 연결되는 정보의 전환 또는 이동 과정이 핵심적인 인간 지능의 행동 원리를 제대로 반영하고 있는지를 정량화하는 문제해결 정보처리 효율 정량화부(1330)를 더 포함할 수 있다.
- [214] 다양한 실시예들에 따르면, 문제해결 정보처리 효율 정량화부(1330)는, 문맥 변화에 따라 문제해결 정책을 변화시키는 인간 모사형 강화학습 모델의 정책 신뢰도 정량화부를 통한 적응 능력과 일반화 능력 검증부를 통한 문제해결을 위해 검증된 일반화 능력과의 연결성 확인을 위해 마르코프 체인(Markov chain)을 이용하여 정량화할 수 있다.
- [215] 다양한 실시예들에 따르면, 문제해결 정보처리 효율 정량화부(1330)는, 문제해결 과정에서 발생하는 과거 에피소드가 강화학습 정책에 반영되는 정보 압축 효율과 강화학습 정책으로부터 도출되는 행동의 최적성 비율을 에피소드-정책-행동으로 이어지는 마르코프 체인상의 상호 정보량(mutual information)을 이용하여 계산할 수 있다.
- [216] 다양한 실시예들에 따르면, 행동의 최적성 비율은 최적의 문제해결을 위해 에피소드 정보를 강화학습 정책결정에 반영하는 정보전달 체계를 나타내는 지표가 될 수 있다.
- [217] 다양한 실시예들에 따르면, 강화학습 모델은, 인간이 학습한 정책 정보를 신뢰성 있게 인코딩하는 모델 기반 제어와 모델 없는(model-free) 제어를 결합한 계산 모델일 수 있다.
- [218]
- [219] 본 문서의 다양한 실시예들은 컴퓨터 장치에 의해 읽을 수 있는 기록 매체(storage medium)에 저장된 하나 이상의 명령들을 포함하는 컴퓨터 프로그램으로서 구현될 수 있다. 예를 들면, 컴퓨터 장치의 프로세서(예: 프로세서(140))는, 기록 매체로부터 저장된 하나 이상의 명령들 중 적어도 하나를 호출하고, 그것을 실행할 수 있다. 이것은 컴퓨터 장치가 호출된 적어도 하나의 명령에 따라 적어도 하나의 기능을 수행하도록 운영되는 것을 가능하게 한다. 하나 이상의 명령들은 컴파일러에 의해 생성된 코드 또는 인터프리터에 의해 실행될 수 있는 코드를 포함할 수 있다. 컴퓨터 장치로 읽을 수 있는 기록 매체는, 비일시적(non-transitory) 기록 매체의 형태로 제공될 수 있다. 여기서, ‘비일시적’은 기록 매체가 실재(tangible)하는 장치이고, 신호(signal)(예:

전자기파)를 포함하지 않는다는 것을 의미할 뿐이며, 이 용어는 데이터가 기록 매체에 반영구적으로 저장되는 경우와 임시적으로 저장되는 경우를 구분하지 않는다.

- [220] 본 문서의 다양한 실시예들 및 이에 사용된 용어들은 본 문서에 기재된 기술을 특정한 실시 형태에 대해 한정하려는 것이 아니며, 해당 실시 예의 다양한 변경, 균등물, 및/또는 대체물을 포함하는 것으로 이해되어야 한다. 도면의 설명과 관련하여, 유사한 구성 요소에 대해서는 유사한 참조 부호가 사용될 수 있다. 단수의 표현은 문맥상 명백하게 다르게 뜻하지 않는 한, 복수의 표현을 포함할 수 있다. 본 문서에서, "A 또는 B", "A 및/또는 B 중 적어도 하나", "A, B 또는 C" 또는 "A, B 및/또는 C 중 적어도 하나" 등의 표현은 함께 나열된 항목들의 모든 가능한 조합을 포함할 수 있다. "제 1", "제 2", "첫째" 또는 "둘째" 등의 표현들은 해당 구성 요소들을, 순서 또는 중요도에 상관없이 수식할 수 있고, 한 구성 요소를 다른 구성 요소와 구분하기 위해 사용될 뿐 해당 구성 요소들을 한정하지 않는다. 어떤(예: 제 1) 구성 요소가 다른(예: 제 2) 구성 요소에 "(기능적으로 또는 통신적으로) 연결되어" 있다거나 "접속되어" 있다고 언급된 때에는, 상기 어떤 구성 요소가 상기 다른 구성 요소에 직접적으로 연결되거나, 다른 구성 요소(예: 제 3 구성 요소)를 통하여 연결될 수 있다.
- [221] 본 문서에서 사용된 용어 "모듈"은 하드웨어, 소프트웨어 또는 펌웨어로 구성된 유닛을 포함하며, 예를 들면, 로직, 논리 블록, 부품, 또는 회로 등의 용어와 상호 호환적으로 사용될 수 있다. 모듈은, 일체로 구성된 부품 또는 하나 또는 그 이상의 기능을 수행하는 최소 단위 또는 그 일부가 될 수 있다. 예를 들면, 모듈은 ASIC(application-specific integrated circuit)으로 구성될 수 있다.
- [222] 다양한 실시예들에 따르면, 기술한 구성 요소들의 각각의 구성 요소(예: 모듈 또는 프로그램)는 단수 또는 복수의 개체를 포함할 수 있다. 다양한 실시예들에 따르면, 기술한 해당 구성 요소들 중 하나 이상의 구성 요소들 또는 동작들이 생략되거나, 또는 하나 이상의 다른 구성 요소들 또는 동작들이 추가될 수 있다. 대체적으로 또는 추가적으로, 복수의 구성 요소들(예: 모듈 또는 프로그램)은 하나의 구성 요소로 통합될 수 있다. 이런 경우, 통합된 구성 요소는 복수의 구성 요소들 각각의 구성 요소의 하나 이상의 기능들을 통합 이전에 복수의 구성 요소들 중 해당 구성 요소에 의해 수행되는 것과 동일 또는 유사하게 수행할 수 있다. 다양한 실시예들에 따르면, 모듈, 프로그램 또는 다른 구성 요소에 의해 수행되는 동작들은 순차적으로, 병렬적으로, 반복적으로, 또는 휴리스틱하게 실행되거나, 동작들 중 하나 이상이 다른 순서로 실행되거나, 생략되거나, 또는 하나 이상의 다른 동작들이 추가될 수 있다.
- [223]

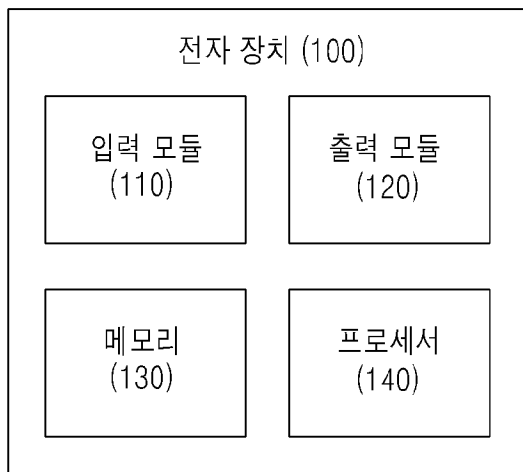
청구범위

- [청구항 1] 전자 장치의 동작 방법에 있어서,
태스크에 대한 인간의 처리 데이터를 기반으로, 1차 모델을 피팅하는 동작;
상기 태스크에 대한 상기 1차 모델의 처리 데이터를 기반으로, 2차 모델을 피팅하는 동작; 및
상기 1차 모델과 상기 2차 모델에 대한 프로파일링을 통해, 상기 2차 모델을 상기 인간의 지능에 대한 이식 모델로 결정하는 동작을 포함하는 방법.
- [청구항 2] 제 1 항에 있어서, 상기 인간의 처리 데이터는,
상기 인간이 상기 태스크를 처리하는 동안 발생하는 행위 데이터 또는 뇌 신호 중 적어도 어느 하나를 포함하는 방법.
- [청구항 3] 제 1 항에 있어서, 상기 이식 모델로 결정하는 동작은,
상기 1차 모델과 상기 2차 모델의 상관도를 검출하는 동작; 및
상기 상관도를 기반으로, 상기 2차 모델을 상기 이식 모델로 결정할지의 여부를 판단하는 동작을 포함하는 방법.
- [청구항 4] 제 3 항에 있어서,
이론적으로 적어도 하나의 환경 변수를 설계하는 동작을 더 포함하고,
상기 1차 모델을 피팅하는 동작은,
상기 환경 변수를 기반으로, 상기 인간의 처리 데이터로부터 상기 1차 모델을 피팅하고,
상기 2차 모델을 피팅하는 동작은,
상기 환경 변수를 기반으로, 상기 1차 모델의 처리 데이터로부터 상기 2차 모델을 피팅하는 방법.
- [청구항 5] 제 4 항에 있어서, 상기 1차 모델을 피팅하는 동작은,
상기 인간의 처리 데이터를 기반으로, 상기 1차 모델을 학습하는 동작을 포함하고,
이로 인해, 상기 1차 모델의 행동 프로파일 또는 적어도 하나의 파라미터 중 적어도 어느 하나가 상기 환경 변수를 기반으로 검출되는 방법.
- [청구항 6] 제 5 항에 있어서, 상기 2차 모델을 피팅하는 동작은,
상기 1차 모델의 처리 데이터를 기반으로, 상기 2차 모델을 학습하는 동작을 포함하고,
이로 인해, 상기 2차 모델의 행동 프로파일 또는 적어도 하나의 파라미터 중 적어도 어느 하나가 상기 환경 변수를 기반으로 검출되는 방법.
- [청구항 7] 제 6 항에 있어서, 상기 상관도를 검출하는 동작은,
상기 1차 모델의 행동 프로파일과 상기 2차 모델의 행동 프로파일을 비교하여, 프로파일 상관도를 검출하는 동작, 또는

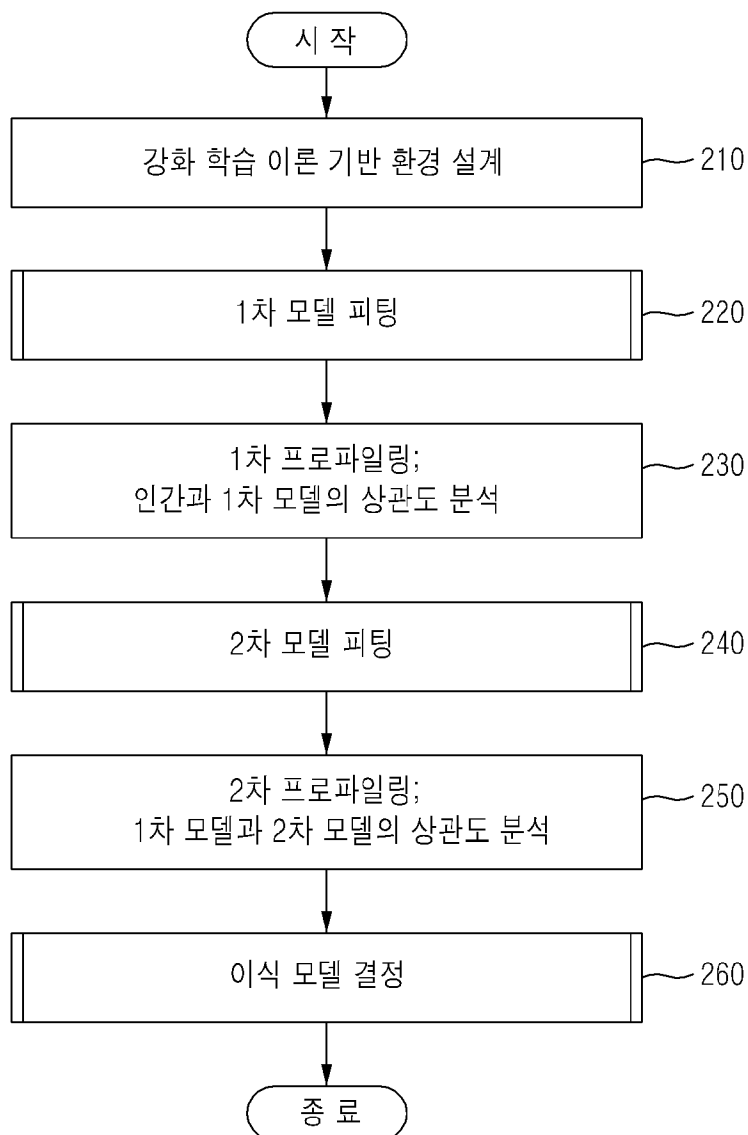
- 상기 1차 모델의 파라미터와 상기 2차 모델의 파라미터를 비교하여,
파라미터 상관도를 검출하는 동작 중 적어도 어느 하나; 및
상기 프로파일 상관도 또는 상기 파라미터 상관도 중 적어도 어느 하나를
기반으로, 상기 상관도를 검출하는 동작을 포함하는 방법.
- [청구항 8] 제 3 항에 있어서, 상기 이식 모델로 결정할 지의 여부를 판단하는 동작은,
상기 상관도가 미리 설정되는 임계값을 초과하면, 상기 2차 모델을 상기
이식 모델로 결정하는 동작을 포함하는 방법.
- [청구항 9] 제 4 항에 있어서, 상기 환경 변수는,
상태-전이 불확실성, 상태-공간 복잡성, 신규성, 상태 예측 오류 또는 보상
예측 오류 중 적어도 어느 하나를 포함하는 방법.
- [청구항 10] 전자 장치에 있어서,
메모리; 및
상기 메모리와 연결되고, 상기 메모리에 저장된 적어도 하나의 명령을
실행하도록 구성된 프로세서를 포함하고,
상기 프로세서는,
태스크에 대한 인간의 처리 데이터를 기반으로, 1차 모델을 피팅하고,
상기 태스크에 대한 상기 1차 모델의 처리 데이터를 기반으로, 2차 모델을
피팅하고,
상기 1차 모델과 상기 2차 모델에 대한 프로파일링을 통해, 상기 2차
모델을 상기 인간의 지능에 대한 이식 모델로 결정하도록 구성되는 장치.
- [청구항 11] 제 10 항에 있어서, 상기 인간의 처리 데이터는,
상기 인간이 상기 태스크를 처리하는 동안 발생하는 행위 데이터 또는 뇌
신호 중 적어도 어느 하나를 포함하는 장치.
- [청구항 12] 제 10 항에 있어서, 상기 프로세서는,
상기 1차 모델과 상기 2차 모델의 상관도를 검출하고,
상기 상관도를 기반으로, 상기 2차 모델을 상기 이식 모델로 결정할 지의
여부를 판단하도록 구성되는 장치.
- [청구항 13] 제 12 항에 있어서, 상기 프로세서는,
이론적으로 적어도 하나의 환경 변수를 설계하고,
상기 환경 변수를 기반으로, 상기 인간의 처리 데이터로부터 상기 1차
모델을 피팅하고,
상기 환경 변수를 기반으로, 상기 1차 모델의 처리 데이터로부터 상기 2차
모델을 피팅하도록 구성되는 장치.
- [청구항 14] 제 13 항에 있어서, 상기 프로세서는,
상기 인간의 처리 데이터를 기반으로, 상기 1차 모델을 학습하도록
구성되고,
이로 인해, 상기 1차 모델의 행동 프로파일 또는 적어도 하나의 파라미터
중 적어도 어느 하나가 상기 환경 변수를 기반으로 검출되는 장치.

- [청구항 15] 제 14 항에 있어서, 상기 프로세서는,
상기 1차 모델의 처리 데이터를 기반으로, 상기 2차 모델을 학습하도록 구성되고,
이로 인해, 상기 2차 모델의 행동 프로파일 또는 적어도 하나의 파라미터 중 적어도 어느 하나가 상기 환경 변수를 기반으로 검출되는 장치.
- [청구항 16] 제 15 항에 있어서, 상기 프로세서는,
상기 1차 모델의 행동 프로파일과 상기 2차 모델의 행동 프로파일을 비교하여, 프로파일 상관도를 검출하고,
상기 1차 모델의 파라미터와 상기 2차 모델의 파라미터를 비교하여, 파라미터 상관도를 검출하고,
상기 프로파일 상관도 또는 상기 파라미터 상관도 중 적어도 어느 하나를 기반으로, 상기 상관도를 검출하도록 구성되는 장치.
- [청구항 17] 제 12 항에 있어서, 상기 프로세서는,
상기 상관도가 미리 설정되는 임계값을 초과하면, 상기 2차 모델을 상기 이식 모델로 결정하도록 구성되는 장치.
- [청구항 18] 제 13 항에 있어서, 상기 환경 변수는,
상태-전이 불확실성, 상태-공간 복잡성, 신규성, 상태 예측 오류 또는 보상 예측 오류 중 적어도 어느 하나를 포함하는 장치.
- [청구항 19] 컴퓨터 장치에 결합되어, 상기 컴퓨터 장치에 의해 판독 가능한 기록 매체에 저장된 컴퓨터 프로그램에 있어서,
태스크에 대한 인간의 처리 데이터를 기반으로, 1차 모델을 피팅하는 동작;
상기 태스크에 대한 상기 1차 모델의 처리 데이터를 기반으로, 2차 모델을 피팅하는 동작; 및
상기 1차 모델과 상기 2차 모델에 대한 프로파일링을 통해, 상기 2차 모델을 상기 인간의 지능에 대한 이식 모델로 결정하는 동작을 실행하기 위한 컴퓨터 프로그램.
- [청구항 20] 제 19 항에 있어서, 상기 이식 모델로 결정하는 동작은,
상기 1차 모델과 상기 2차 모델의 상관도를 검출하는 동작; 및
상기 상관도를 기반으로, 상기 2차 모델을 상기 이식 모델로 결정할 지의 여부를 판단하는 동작을 포함하는 컴퓨터 프로그램.

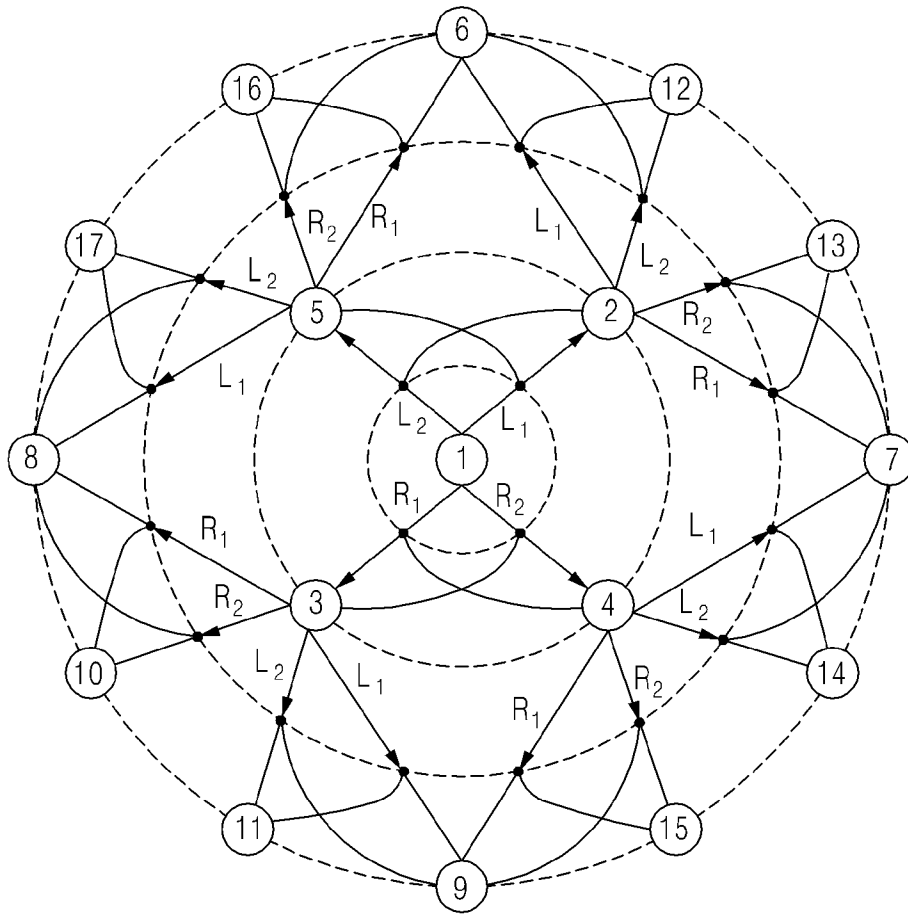
[도1]



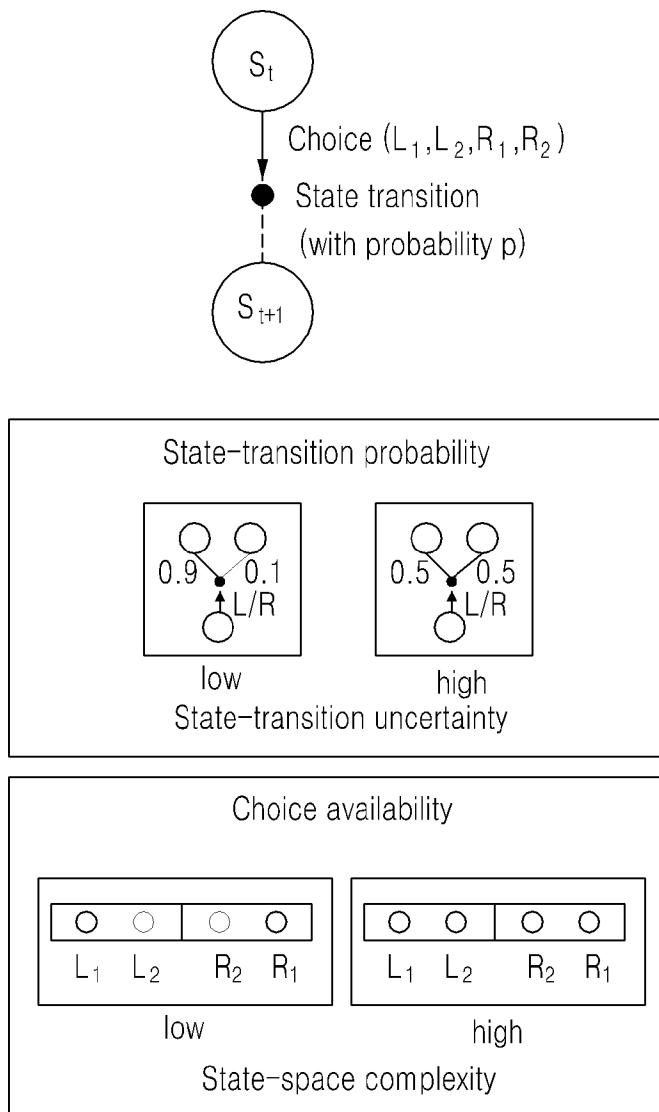
[도2]



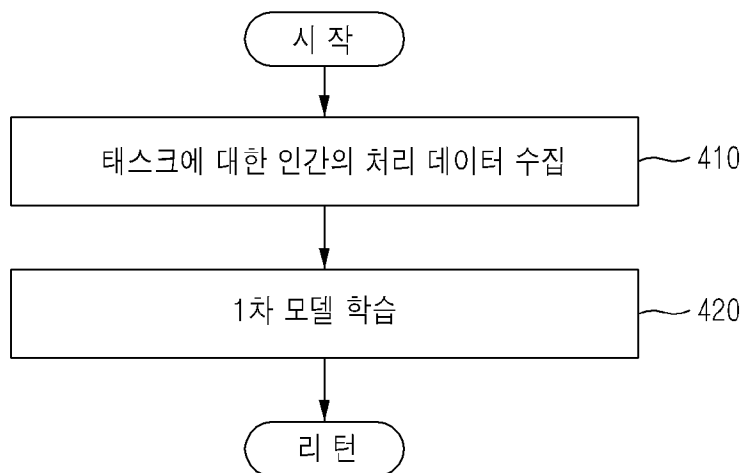
[도3a]



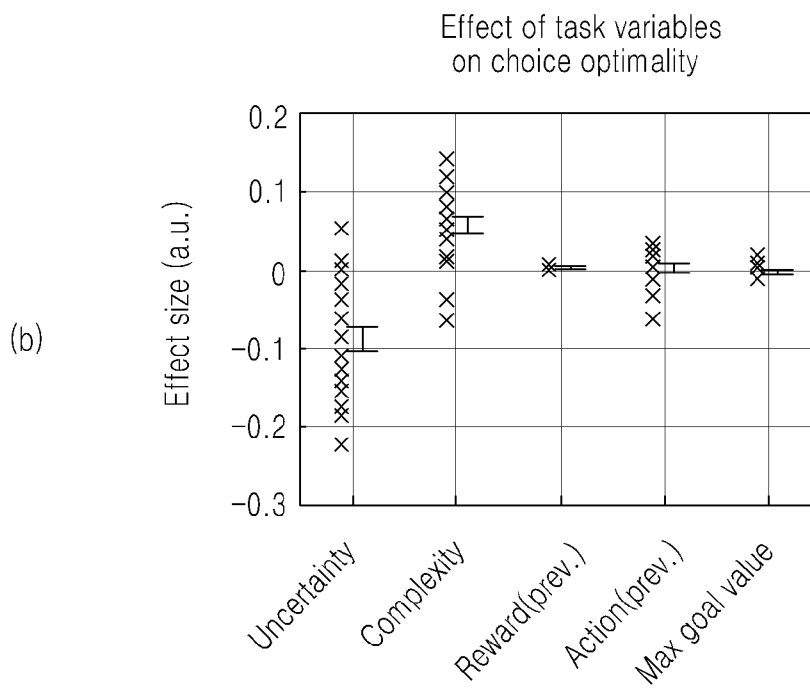
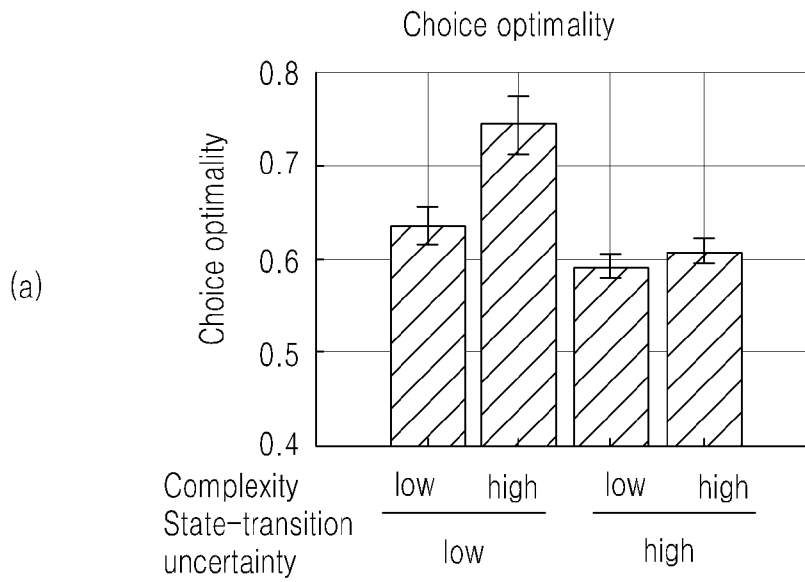
[도3b]



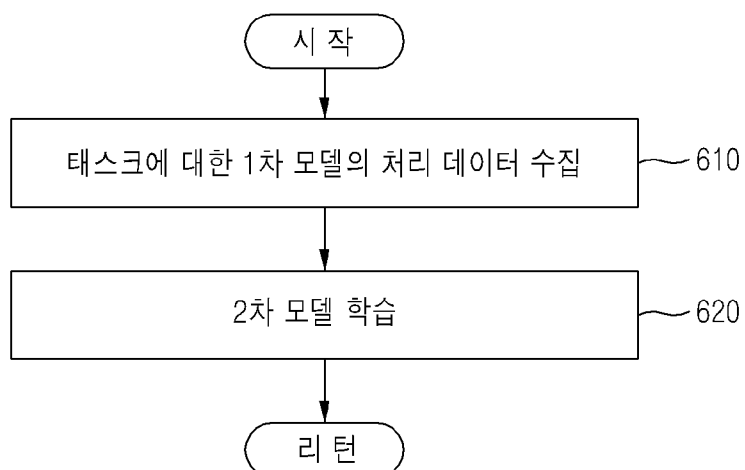
[도4]



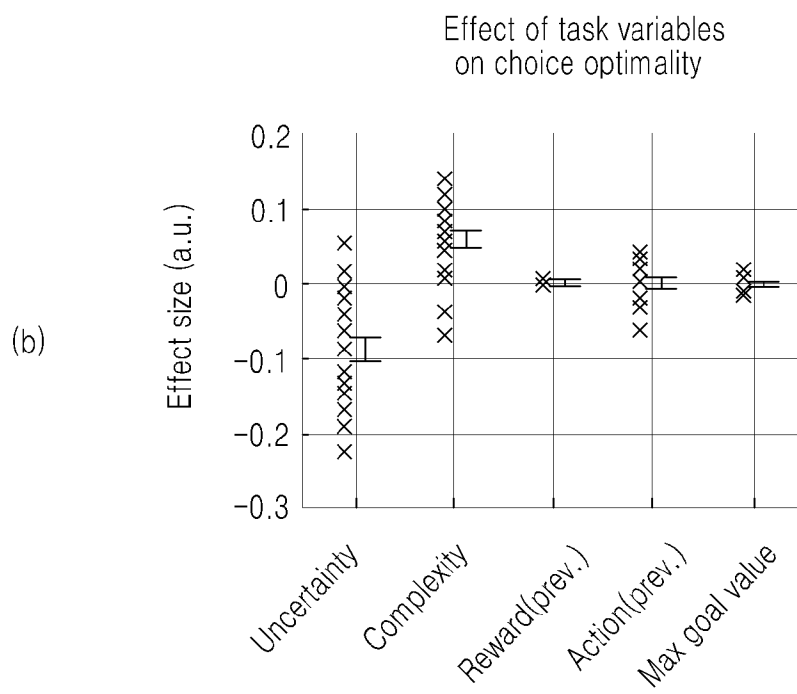
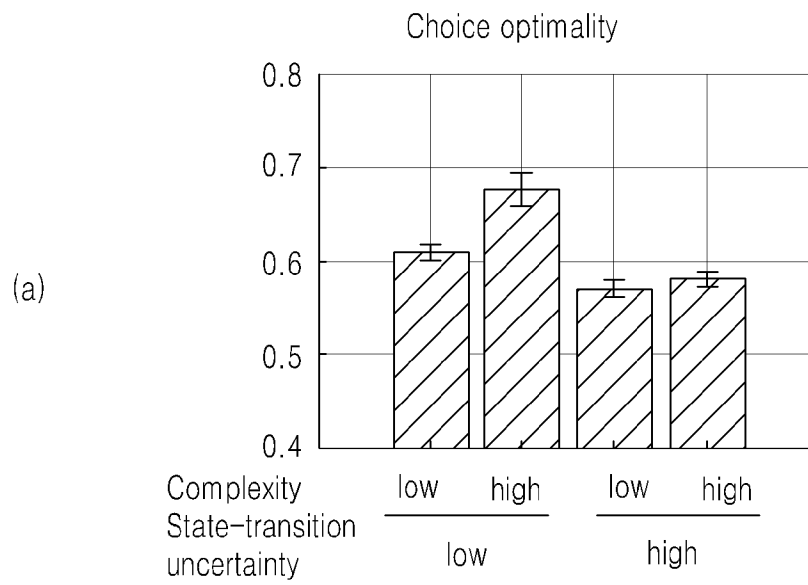
[도5]



[도6]



[도7]



[도8]

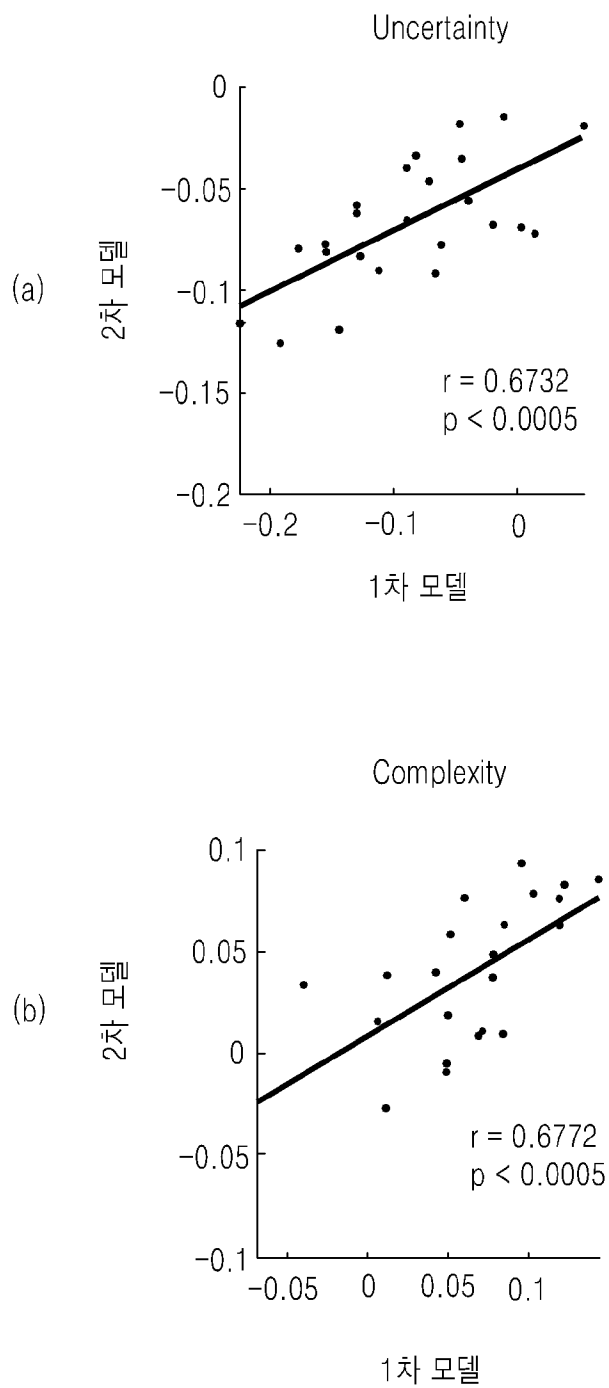
(a)

Effect (repeated measure ANOVA)Uncertainty : $p < 0.001$ Complexity : $p < 0.001$ Uncertainty x complexity : $p < 0.001$

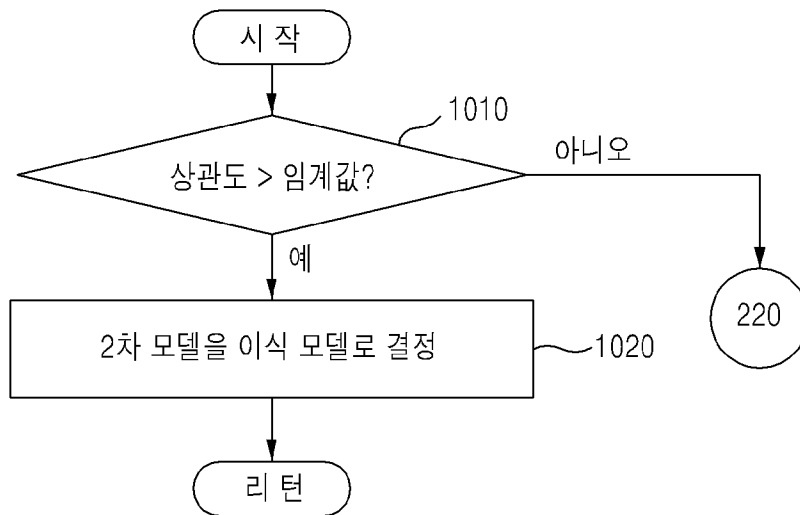
(b)

Effect (repeated measure ANOVA)Uncertainty : $p < 0.001$ Complexity : $p < 0.001$ Uncertainty x complexity : $p < 0.001$

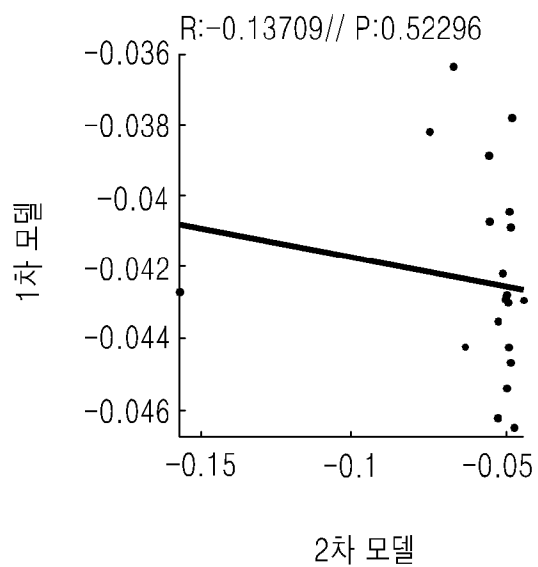
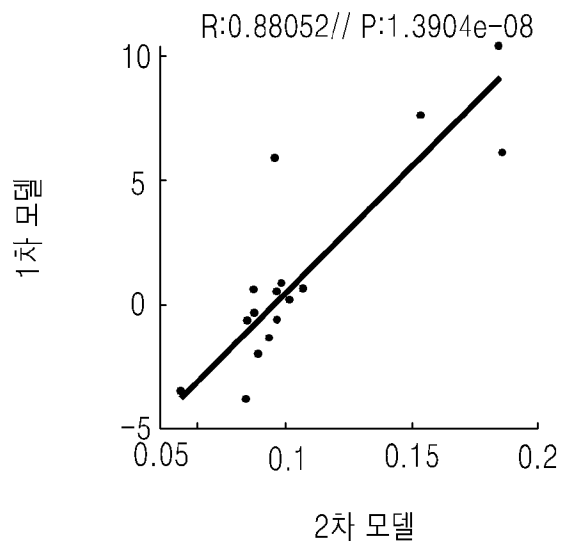
[도9]



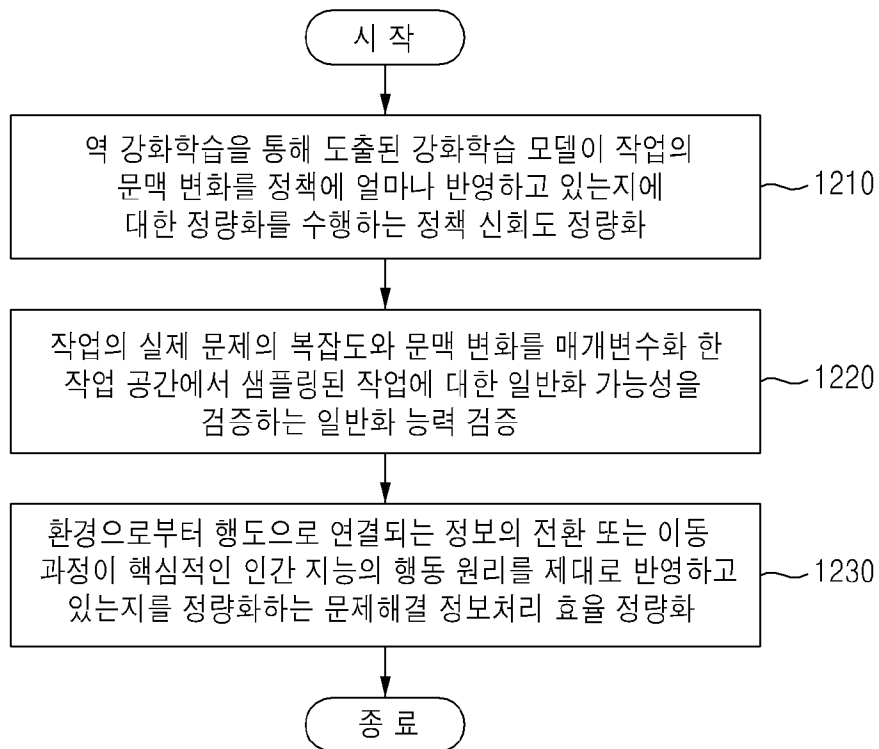
[도10]



[도11]

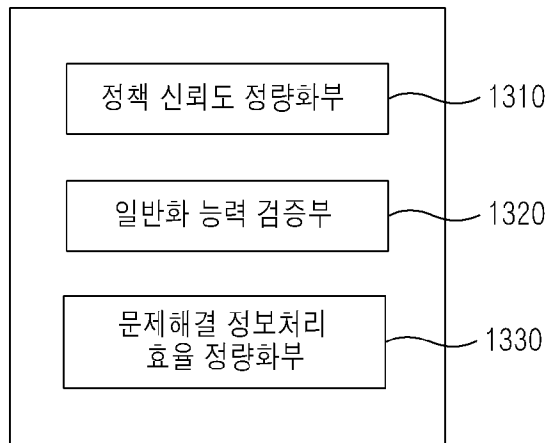


[도12]

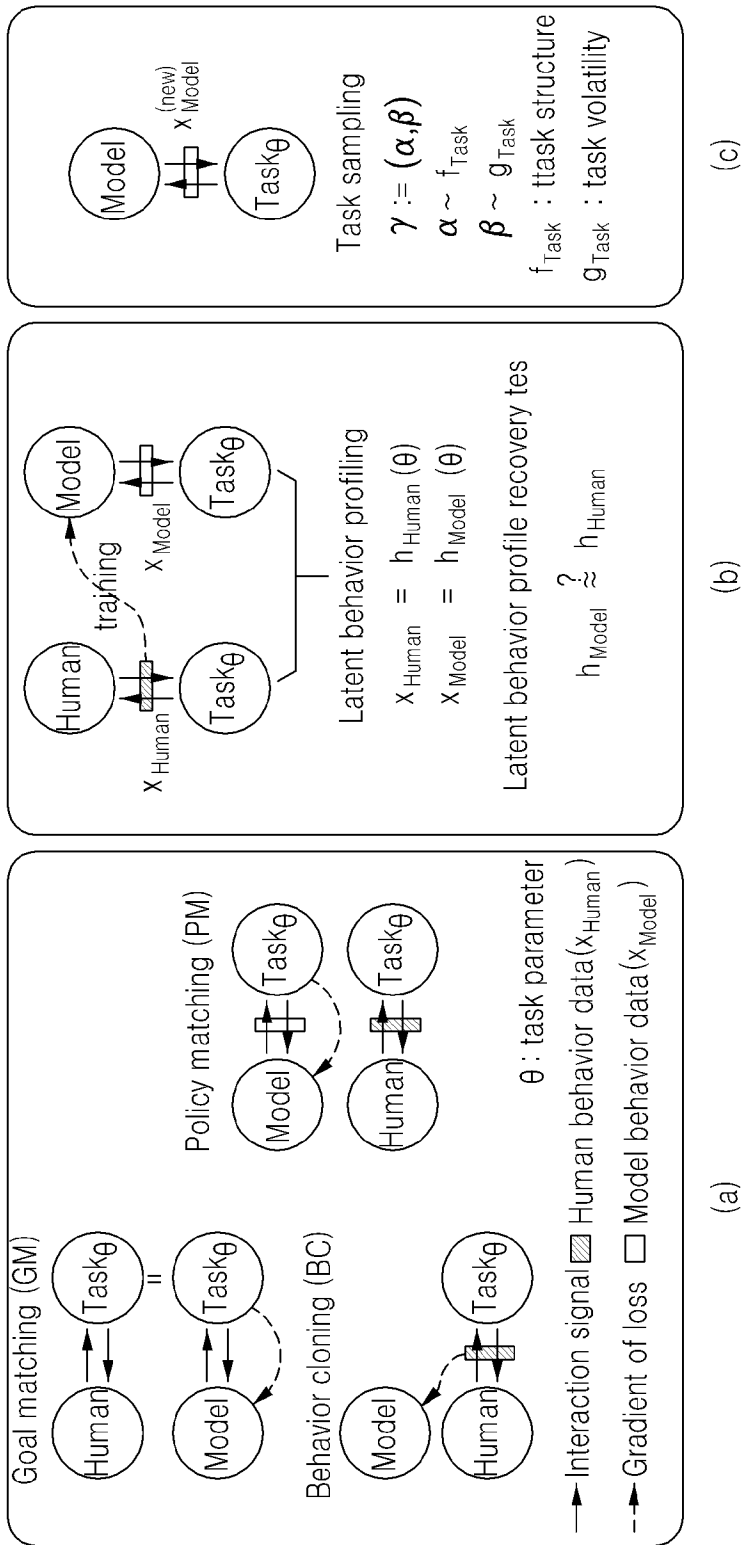


[도13]

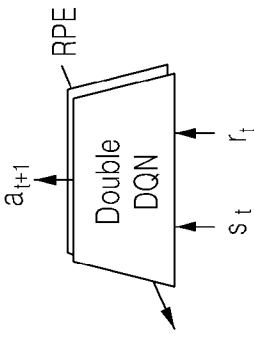
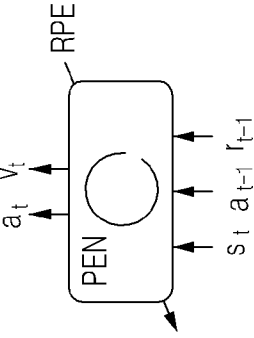
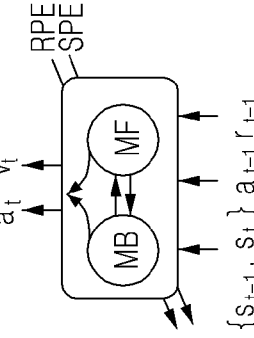
1300



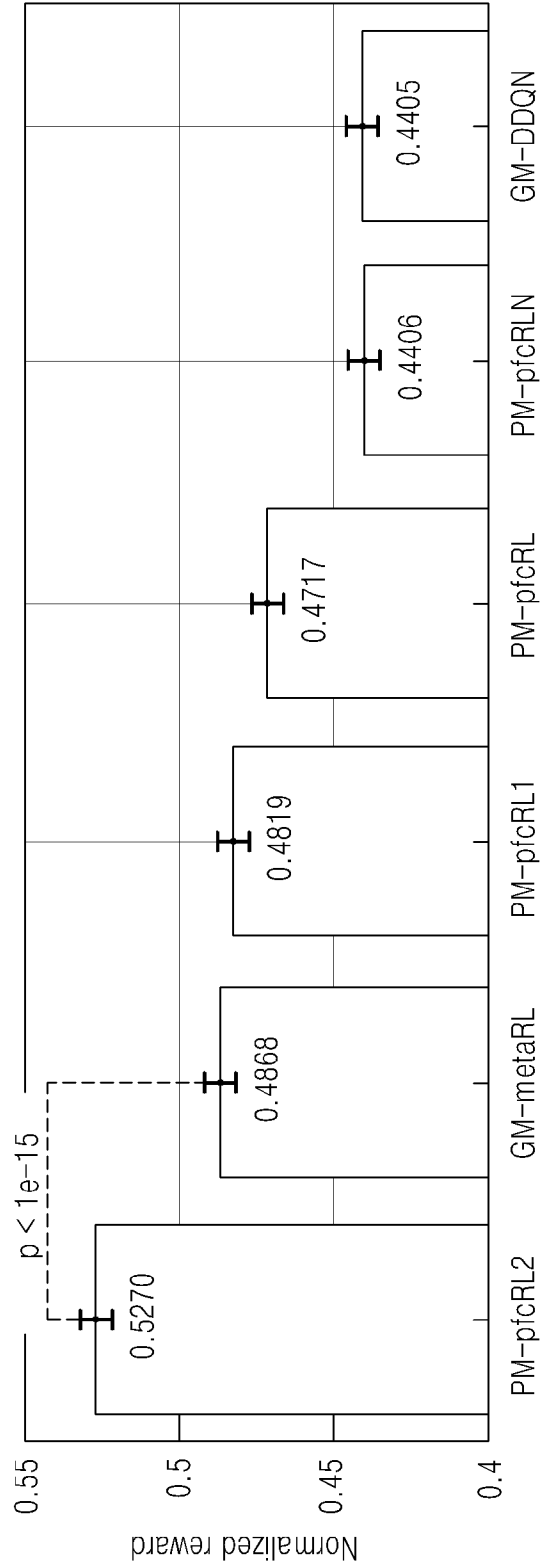
[도 14]



[도15]

Category	Deep RL	Meta RL	Prefrontal RL
Architecture			
Learning strategy	Model-free	Model-free + Model-based	Model-free x Model-based
Training method	Goal matching (GM) Policy matching (PM)	Goal matching (GM) Policy matching (PM)	Policy matching (PM)
Model name	GM-DDQN, PM-DDQN	GM-metaRL, PM-metaRL	PM-pfcRL1, PM-pfcRL2

[도17]



INTERNATIONAL SEARCH REPORT

International application No.

PCT/KR2020/018025

A. CLASSIFICATION OF SUBJECT MATTER G06N 20/00(2019.01)i; G06N 3/00(2006.01)i According to International Patent Classification (IPC) or to both national classification and IPC		
B. FIELDS SEARCHED Minimum documentation searched (classification system followed by classification symbols) G06N 20/00(2019.01); G06F 15/18(2006.01); G06F 16/35(2019.01); G06N 3/04(2006.01); G06N 3/08(2006.01); G06N 5/00(2006.01); G06N 5/04(2006.01) Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched Korean utility models and applications for utility models: IPC as above Japanese utility models and applications for utility models: IPC as above Electronic data base consulted during the international search (name of data base and, where practicable, search terms used) eKOMPASS (KIPO internal) & keywords: 인간지능 이식 모델(human intelligence transplanting model), 정밀 행동 프로파일링(behavioral behavior profiling), 행위 데이터(behavioral data), 기능적 MRI(fMRI; functional magnetic resonance imaging)		
C. DOCUMENTS CONSIDERED TO BE RELEVANT		
Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y	KR 10-2020-0017595 A (KOREA ADVANCED INSTITUTE OF SCIENCE AND TECHNOLOGY) 19 February 2020 (2020-02-19) See paragraphs [0030]-[0033] and [0064]-[0069]; and figure 2.	1-20
Y	KIM, Dongjae et al. Task complexity interacts with state-space uncertainty in the arbitration between model-based and model-free learning. Article [online]. NATURE COMMUNICATIONS. PP. 1-14, 16 December 2019. [Retrieved on 22 February 2021]. Retrieved from <URL: https://www.nature.com/articles/s41467-019-13632-1 >. See pages 5-13; and figures 4-6e.	1-20
A	US 9280745 B1 (APPLIED UNDERWRITERS, INC.) 08 March 2016 (2016-03-08) See columns 1-2 and 7; and figure 2.	1-20
A	KR 10-2013-0071734 A (KOREA ADVANCED INSTITUTE OF SCIENCE AND TECHNOLOGY) 01 July 2013 (2013-07-01) See paragraphs [0002], [0010], [0022]-[0023] and [0027]; and figure 1.	1-20
☒ Further documents are listed in the continuation of Box C. ☒ See patent family annex.		
* Special categories of cited documents: "A" document defining the general state of the art which is not considered to be of particular relevance "D" document cited by the applicant in the international application "E" earlier application or patent but published on or after the international filing date "L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified) "O" document referring to an oral disclosure, use, exhibition or other means "P" document published prior to the international filing date but later than the priority date claimed "T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention "X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone "Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art "&" document member of the same patent family		
Date of the actual completion of the international search 05 March 2021		Date of mailing of the international search report 05 March 2021
Name and mailing address of the ISA/KR Korean Intellectual Property Office Government Complex-Daejeon Building 4, 189 Cheongsaro, Seo-gu, Daejeon 35208 Facsimile No. +82-42-481-8578		Authorized officer Telephone No.

INTERNATIONAL SEARCH REPORT

International application No.

PCT/KR2020/018025

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	KR 10-2019-0106861 A (LG ELECTRONICS INC.) 18 September 2019 (2019-09-18) See paragraphs [0023]-[0026] and [0185]-[0186]; and figure 4.	1-20

INTERNATIONAL SEARCH REPORT
Information on patent family members

International application No.

PCT/KR2020/018025

Patent document cited in search report			Publication date (day/month/year)	Patent family member(s)			Publication date (day/month/year)
KR	10-2020-0017595	A	19 February 2020	KR	10-2132529	B1	09 July 2020
				US	2020-0043358	A1	06 February 2020
US	9280745	B1	08 March 2016	None			
KR	10-2013-0071734	A	01 July 2013	None			
KR	10-2019-0106861	A	18 September 2019	US	2020-0034661	A1	30 January 2020

A. 발명이 속하는 기술분류(국제특허분류(IPC)) G06N 20/00(2019.01)i; G06N 3/00(2006.01)i		
B. 조사된 분야 조사된 최소문헌(국제특허분류를 기재) G06N 20/00(2019.01); G06F 15/18(2006.01); G06F 16/35(2019.01); G06N 3/04(2006.01); G06N 3/08(2006.01); G06N 5/00(2006.01); G06N 5/04(2006.01) 조사된 기술분야에 속하는 최소문헌 이외의 문헌 한국등록실용신안공보 및 한국공개실용신안공보: 조사된 최소문헌란에 기재된 IPC 일본등록실용신안공보 및 일본공개실용신안공보: 조사된 최소문헌란에 기재된 IPC 국제조사에 이용된 전산 데이터베이스(데이터베이스의 명칭 및 검색어(해당하는 경우)) eKOMPASS(특허청 내부 검색시스템) & 키워드: 인간지능 이식 모델(human intelligence transplanting model), 정밀 행동 프로파일링(behavioral behavior profiling), 행위 데이터(behavioral data), 기능적 MRI(fMRI; functional magnetic resonance imaging)		
C. 관련 문헌		
카테고리*	인용문헌명 및 관련 구절(해당하는 경우)의 기재	관련 청구항
Y	KR 10-2020-0017595 A (한국과학기술원) 2020.02.19 단락 [0030]-[0033], [0064]-[0069]; 및 도면 2	1-20
Y	DONGJAE KIM et al, Task complexity interacts with state-space uncertainty in the arbitration between model-based and model-free learning, Article [online], NATURE COMMUNICATIONS, PP. 1-14, 2019.12.16 [검색일: 2021-02-22]. 출처 <URL: https://www.nature.com/articles/s41467-019-13632-1 > 페이지 5-13; 및 도면 4-6e	1-20
A	US 9280745 B1 (APPLIED UNDERWRITERS, INC.) 2016.03.08 컬럼 1-2, 7; 및 도면 2	1-20
A	KR 10-2013-0071734 A (한국과학기술원) 2013.07.01 단락 [0002], [0010], [0022]-[0023], [0027]; 및 도면 1	1-20
☒ 추가 문헌이 C(계속)에 기재되어 있습니다. ☒ 대응특허에 관한 별지를 참조하십시오.		
* 인용된 문헌의 특별 카테고리: “A” 특별히 관련이 없는 것으로 보이는 일반적인 기술수준을 정의한 문헌 “D” 본 국제출원에서 출원인이 인용한 문헌 “E” 국제출원일보다 빠른 출원일 또는 우선일을 가지나 국제출원일 이후에 공개된 선출원 또는 특허 문헌 “L” 우선권 주장에 의문을 제기하는 문헌 또는 다른 인용문헌의 공개일 또는 다른 특별한 이유(이유를 명시)를 밝히기 위하여 인용된 문헌 “O” 구두 개시, 사용, 전시 또는 기타 수단을 언급하고 있는 문헌 “P” 우선일 이후에 공개되었으나 국제출원일 이전에 공개된 문헌 “T” 국제출원일 또는 우선일 후에 공개된 문헌으로, 출원과 상충하지 않으며 발명의 기초가 되는 원리나 이론을 이해하기 위해 인용된 문헌 “X” 특별한 관련이 있는 문헌. 해당 문헌 하나만으로 청구된 발명의 신규성 또는 진보성이 없는 것으로 본다. “Y” 특별한 관련이 있는 문헌. 해당 문헌이 하나 이상의 다른 문헌과 조합하는 경우로 그 조합이 당업자에게 자명한 경우 청구된 발명은 진보성이 없는 것으로 본다. “&” 동일한 대응특허문헌에 속하는 문헌		
국제조사의 실제 완료일 2021년03월05일(05.03.2021)		국제조사보고서 발송일 2021년03월05일(05.03.2021)
ISA/KR의 명칭 및 우편주소 대한민국 특허청 (35208) 대전광역시 서구 청사로 189, 4동 (둔산동, 정부대전청사) 팩스 번호 +82-42-481-8578		심사관 양정록 전화번호 +82-42-481-5709

C. 관련 문헌		
카테고리*	인용문헌명 및 관련 구절(해당하는 경우)의 기재	관련 청구항
A	KR 10-2019-0106861 A (엘지전자 주식회사) 2019.09.18 단락 [0023]-[0026], [0185]-[0186]; 및 도면 4	1-20

국 제 조 사 보 고 서
대응특허에 관한 정보

국제출원번호

PCT/KR2020/018025

국제조사보고서에서 인용된 특허문헌	공개일	대응특허문헌	공개일
KR 10-2020-0017595 A	2020/02/19	KR 10-2132529 B1	2020/07/09
		US 2020-0043358 A1	2020/02/06
US 9280745 B1	2016/03/08	없음	
KR 10-2013-0071734 A	2013/07/01	없음	
KR 10-2019-0106861 A	2019/09/18	US 2020-0034661 A1	2020/01/30