



(19) 대한민국특허청(KR)
(12) 공개특허공보(A)

(11) 공개번호 10-2025-0018622
(43) 공개일자 2025년02월07일

(51) 국제특허분류(Int. Cl.)
A61B 5/00 (2021.01) A61B 5/369 (2025.01)
G06N 20/00 (2019.01) G16H 50/20 (2018.01)
(52) CPC특허분류
A61B 5/7267 (2013.01)
A61B 5/369 (2025.01)
(21) 출원번호 10-2023-0099442
(22) 출원일자 2023년07월31일
심사청구일자 2023년07월31일

(71) 출원인
고려대학교 산학협력단
서울특별시 성북구 안암로 145, 고려대학교 (안암동5가)
(72) 발명자
김태의
서울특별시 성북구 동소문로44길 17, 101동 1805호
신동희
대전광역시 서구 도마로7번길 3
(74) 대리인
김준석, 박민욱

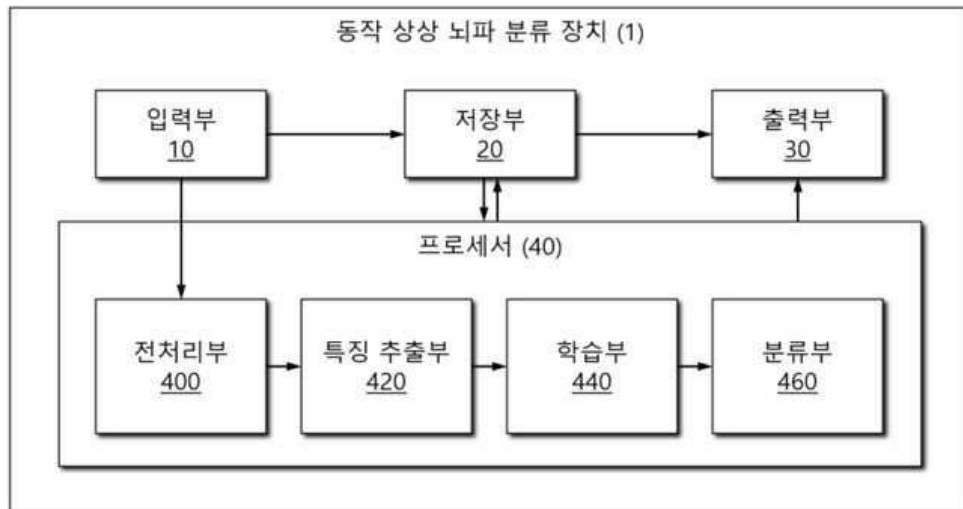
전체 청구항 수 : 총 10 항

(54) 발명의 명칭 멀티에이전트 강화학습을 활용한 동작 상상 뇌파 분류 장치 및 방법

(57) 요약

본 발명은 멀티에이전트 강화학습을 활용한 동작 상상 뇌파 분류 장치에 관한 것으로, 뇌파 데이터를 전처리하는 전처리부, 전처리된 뇌파 데이터의 복수의 도메인에서 의미 있는 특징을 전체 특징 집합으로 추출하는 특징 추출부, 복수의 에이전트를 활용하여 각 에이전트가 복수의 도메인에 대해 관찰 및 액션을 수행하고 피드백 하는 멀티에이전트 강화학습을 통해 전체 특징 집합 중 동작 상상에 의미 있는 주요 특징을 추출하는 학습부 및 주요 특징을 통해 뇌파 데이터로부터 사용자의 움직임 의도를 판단함으로써 동작 상상을 분류하는 분류부를 포함하는 동작 상상 뇌파 분류 장치.

대표도 - 도1



(52) CPC특허분류

G06N 20/00 (2021.08)

G16H 50/20 (2018.01)

이 발명을 지원한 국가연구개발사업

과제고유번호	1711193973
과제번호	2019-0-00079-005
부처명	과학기술정보통신부
과제관리(전문)기관명	정보통신기획평가원
연구사업명	정보통신방송혁신인재양성
연구과제명	인공지능대학원지원(고려대학교)
기 여 율	100/100
과제수행기관명	고려대학교산학협력단
연구기간	2023.01.01 ~ 2023.12.31

명세서

청구범위

청구항 1

멀티에이전트 강화학습을 활용한 동작 상상 너파 분류 장치에 관한 것으로,

너파 데이터를 전처리하는 전처리부;

전처리된 상기 너파 데이터의 복수의 도메인에서 의미 있는 특징을 전체 특징 집합으로 추출하는 특징 추출부;

복수의 에이전트를 활용하여 각 에이전트가 상기 복수의 도메인에 대해 관찰 및 액션을 수행하고 피드백 하는 멀티에이전트 강화학습을 통해 상기 전체 특징 집합 중 동작 상상에 의미 있는 주요 특징을 추출하는 학습부; 및

상기 주요 특징을 통해 상기 너파 데이터로부터 사용자의 움직임 의도를 판단함으로써 동작 상상을 분류하는 분류부; 를 포함하는

동작 상상 너파 분류 장치.

청구항 2

제1항에 있어서,

상기 학습부는,

공간-스펙트럼 에이전트를 통해 상기 전체 특징 집합의 공간-스펙트럼 도메인 관찰을 수행하고, 정책을 수립하고, 액션을 결정 및 수행하는 공간-스펙트럼 특징 모듈;

시간 에이전트를 통해 상기 전체 특징 집합의 시간 도메인 관찰을 수행하고, 정책을 수립하고, 액션을 결정 및 수행하는 시간 특징 모듈; 및

상기 공간-스펙트럼 에이전트 및 상기 시간 에이전트의 상기 액션 및 이에 따른 보상에 관한 정보를 받고, 상기 공간-스펙트럼 에이전트 및 상기 시간 에이전트와 정책 및 가치를 교환하며 피드백을 수행하는 중앙 집중형 크리틱 네트워크 모듈; 을 포함하는

동작 상상 너파 분류 장치.

청구항 3

제2항에 있어서,

상기 중앙 집중형 크리틱 네트워크 모듈은,

$$Q(s_t, \mathbf{a}_t) = \mathbb{E}_{(s_{t+1:T}, \mathbf{a}_{t+1:T})} [\mathcal{R}_t | s_t, \mathbf{a}_t]$$

(여기서 s_t 는 현재 스텝에서 주어진 특징 전체 집합에 대한 상태, $\mathbf{a}_t$ 는 공간-스펙트럼 에이전트와 시간 에이전트가 수행한 액션을 합친 공동 액션, $\mathcal{R}_t$ 는 현재 상태에서 공동 액션을 수행하였을 때 받게 될 보상의 총합, $\mathbb{E}$ 는 확률 변수에 대한 평균 값을 의미한다.)

의 공동 액션 가치 함수에 기반하여 상기 피드백을 수행하는 것을 특징으로 하는

동작 상상 너파 분류 장치.

청구항 4

제3항에 있어서,

상기 중앙 집중형 크리틱 네트워크 모듈은,

$$A^i(s_t, \mathbf{a}_t) = Q(s_t, \mathbf{a}_t; \psi) - \sum_{\tilde{\mathbf{a}}_t^i} \pi_t^i(\tilde{\mathbf{a}}_t^i | o_t^i; \theta^i) Q(s_t, (\mathbf{a}_t^{-i}, \tilde{\mathbf{a}}_t^i); \psi)$$

(여기서 ψ 는 업데이트가 진행되는 중앙집중형 크리틱의 파라미터, π_t^i 는 해당 에이전트 i 가 가지고 있는 정책, $\tilde{\mathbf{a}}_t^i$ 는 해당 에이전트 i 가 선택할 수 있는 다른 종류의 액션, o_t^i 는 에이전트 i 의 관찰, θ^i 는 에이전트 i 의 파라미터, $\mathbf{a}_t^{-i}$ 는 해당 에이전트 i 를 제외한 다른 에이전트의 모든 액션을 고정시키는 것을 의미한다.)

의 가치 함수를 상기 공간-스펙트럼 에이전트 및 상기 시간 에이전트에 할당되도록 하는 것을 특징으로 하는 동작 상상 뇌파 분류 장치.

청구항 5

제4항에 있어서,

상기 중앙 집중형 크리틱 네트워크 모듈은,

상기 가치 함수를 최대화하는 방향으로 파라미터 θ^i 를 업데이트하는 것을 특징으로 하는

동작 상상 뇌파 분류 장치.

청구항 6

멀티에이전트 강화학습을 활용한 동작 상상 뇌파 분류 방법에 관한 것으로,

뇌파 데이터를 전처리하는 전처리 단계;

전처리된 상기 뇌파 데이터의 복수의 도메인에서 의미 있는 특징을 전체 특징 집합으로 추출하는 단계;

복수의 에이전트를 활용하여 각 에이전트가 상기 복수의 도메인에 대해 관찰 및 액션을 수행하고 피드백 하는 멀티에이전트 강화학습을 통해 상기 전체 특징 집합 중 동작 상상에 의미 있는 주요 특징을 추출하는 단계;

상기 주요 특징을 통해 상기 뇌파 데이터로부터 사용자의 움직임 의도를 판단함으로써 동작 상상을 분류하는 단계;를 포함하는

동작 상상 뇌파 분류 방법.

청구항 7

제6항에 있어서,

상기 주요 특징을 추출하는 단계는,

공간-스펙트럼 에이전트를 통해 상기 전체 특징 집합의 공간-스펙트럼 도메인 관찰을 수행하고, 정책을 수립하고, 액션을 결정 및 수행하는 공간-스펙트럼 특징 선택 단계;

시간 에이전트를 통해 상기 전체 특징 집합의 시간 도메인 관찰을 수행하고, 정책을 수립하고, 액션을 결정 및 수행하는 시간 특징 선택 단계; 및

상기 공간-스펙트럼 에이전트 및 상기 시간 에이전트의 상기 액션 및 이에 따른 보상에 관한 정보를 받고, 상기 공간-스펙트럼 에이전트 및 상기 시간 에이전트와 정책 및 가치를 교환하며 피드백을 수행하는 중앙 집중형 크리틱 단계;를 포함하는

동작 상상 뇌파 분류 방법.

청구항 8

제7항에 있어서,

상기 중앙 집중형 크리틱 단계는,

$$Q(s_t, \mathbf{a}_t) = \mathbb{E}_{(s_{t+1:T}, \mathbf{a}_{t+1:T})}[\mathcal{R}_t | s_t, \mathbf{a}_t]$$

(여기서 S_t 는 현재 스텝에서 주어진 특징 전체 집합에 대한 상태, $\mathbf{a}_t$ 는 공간-스펙트럼 에이전트와 시간 에이전트가 수행한 액션을 합친 공동 액션, R_t 는 현재 상태에서 공동 액션을 수행하였을 때 받게 될 보상의 총합, $\mathbb{E}$ 는 확률 변수에 대한 평균 값을 의미한다.)

의 공동 액션 가치 함수에 기반하여 피드백을 수행하는 것을 특징으로 하는 동작 상상 뇌파 분류 방법.

청구항 9

제8항에 있어서,

상기 중앙 집중형 크리틱 단계는,

$$A^i(s_t, \mathbf{a}_t) = Q(s_t, \mathbf{a}_t; \psi) - \sum_{\tilde{\mathbf{a}}_t^i} \pi_t^i(\tilde{\mathbf{a}}_t^i | o_t^i; \theta^i) Q(s_t, (\mathbf{a}_t^{-i}, \tilde{\mathbf{a}}_t^i); \psi)$$

(여기서 ψ 는 업데이트가 진행되는 중앙집중형 크리틱의 파라미터, π_t^i 는 해당 에이전트 i 가 가지고 있는 정책, $\tilde{\mathbf{a}}_t^i$ 는 해당 에이전트 i 가 선택할 수 있는 다른 종류의 액션, o_t^i 는 에이전트 i 의 관찰, θ^i 는 에이전트 i 의 파라미터, $\mathbf{a}_t^{-i}$ 는 해당 에이전트 i 를 제외한 다른 에이전트의 모든 액션을 고정시키는 것을 의미한다.)

의 가치 함수를 상기 공간-스펙트럼 에이전트 및 상기 시간 에이전트에 할당되도록 하는 것을 특징으로 하는 동작 상상 뇌파 분류 방법.

청구항 10

제9항에 있어서,

상기 중앙 집중형 크리틱 단계는,

상기 가치 함수를 최대화하는 방향으로 파라미터 θ^i 를 업데이트하는 것을 특징으로 하는

동작 상상 뇌파 분류 방법.

발명의 설명

기술 분야

[0001] 동작 상상 뇌파 분류 장치 및 방법에 관한 것으로, 보다 상세하게는 뇌파 데이터 내 동작 상상 예측에 유용한 특징을 추출하기 위해 멀티에이전트 강화학습 기법을 활용한 동작 상상 뇌파 분류 장치 및 방법이다.

배경 기술

[0002] 뇌-컴퓨터 인터페이스(BCI)는 인간의 두뇌와 컴퓨터 또는 외부 기기를 직접 연결함으로써 뇌파를 통해 컴퓨터 또는 외부 기기를 제어하는 인터페이스 기술을 포괄하는 개념이다. 뇌-컴퓨터 인터페이스는 동작 상상(motor imagery, MI)과 관련하여 다양한 응용이 가능하다. 동작 상상은 개인이 실제로 움직이지 않고도 마음속에서 운동을 실행하는 것을 의미하는데, 이러한 동작 상상 활동은 뇌 활동에서 특정한 신호 패턴을 생성하게 된다. BCI는 이러한 신호 패턴을 감지하고 해석하여 컴퓨터나 외부 기기와 상호 작용하는 데 활용된다. 예를 들어, 사용자가 손을 움직이는 것을 상상한다면, 뇌파의 특정 주파수 밴드에서 증가 또는 감소한 신호 패턴이 나타날 수 있다. 그러므로, 이 특정 패턴을 분석하여 사용자의 동적 의도를 사전에 예측하고, 이를 로봇이나 컴퓨터 조작에 적용시킬 수 있다.

[0003] 최근 몇 년 동안 연구자들은 사용자의 동적 의도를 예측하기 위해, 딥러닝 기술을 활용하여 뇌파 신호 데이터에

서 유용한 특징과 패턴을 추출하는 데 집중해왔다. 뇌파 신호 데이터는 시간, 주파수, 공간 등 여러 도메인 측면에서 관찰되며, 이러한 도메인들은 서로 연결되어 종합적인 정보를 제공할 수 있다. 예를 들어, 시간 도메인에서 추출된 특징은 사용자의 의도가 시간적으로 어떻게 변화하는지를 파악하는 데에 도움을 줄 수 있으며, 공간 도메인에서 추출된 특징은 사용자의 의도가 어떤 뇌 영역과 연관되어 있는지를 이해하는 데에 도움을 줄 수 있다.

[0004] 그러나, 이렇게 추출된 특징들은 여러 도메인을 따라 복잡하게 연결되어 있어서 유용한 정보를 포함할 수도 있지만, 동시에 노이즈로 작용할 수도 있다. 최근 여러 연구가 진행됨에 따라 뇌파 신호 데이터 내에서 특징 선택 프로세스를 실행하는 다양한 검색 및 최적화 기반 방법론들이 등장하였지만, 선택된 특징들이 동작 상상과 관련이 있으며 유용성이 있는지 여부를 명확히 판단하기 위해 필요한 도메인적인 지식이 여전히 부족한 상태이다. 이로 인해 검색 및 최적화 기반 방법론을 활용하여 선택된 특징들의 신뢰성을 평가하는 것은 어려운 상황이다.

선행기술문헌

특허문헌

[0005] (특허문헌 0001) 대한민국 공개특허공보 제10-2021-0154759호("사용자의 의도를 고려하여 특정 채널을 맞춤형으로 선택하는 뇌 컴퓨터 인터페이스 장치 및 그 동작 방법", 고려대학교 산학협력단, 2021.12.21)

발명의 내용

해결하려는 과제

[0006] 본 발명은 멀티에이전트 강화학습 기법을 활용함으로써 뇌파 신호 데이터로부터 추출된 특징들 중 동적 상상 분류에 중요한 의미를 갖는 주요 특징들만을 효과적으로 선택하여 사용자의 동적 의도를 보다 정확히 예측할 수 있는 동적 상상 뇌파 분류 장치 및 방법을 제공하고자 한다.

과제의 해결 수단

[0007] 본 발명의 일 실시예는 멀티에이전트 강화학습을 활용한 동작 상상 뇌파 분류 장치에 관한 것으로, 뇌파 데이터를 전처리하는 전처리부, 전처리된 뇌파 데이터의 복수의 도메인에서 의미 있는 특징을 전체 특징 집합으로 추출하는 특징 추출부, 복수의 에이전트를 활용하여 각 에이전트가 복수의 도메인에 대해 관찰 및 액션을 수행하고 피드백 하는 멀티에이전트 강화학습을 통해 전체 특징 집합 중 동작 상상에 의미 있는 주요 특징을 추출하는 학습부 및 주요 특징을 통해 상기 뇌파 데이터로부터 사용자의 움직임 의도를 판단함으로써 동작 상상을 분류하는 분류부를 포함할 수 있다.

[0008] 또한, 일 실시예에 따라서, 학습부는, 공간-스펙트럼 에이전트를 통해 전체 특징 집합의 공간-스펙트럼 도메인 관찰을 수행하고, 정책을 수립하고, 액션을 결정 및 수행하는 공간-스펙트럼 특징 모듈, 시간 에이전트를 통해 전체 특징 집합의 시간 도메인 관찰을 수행하고, 정책을 수립하고, 액션을 결정 및 수행하는 시간 특징 모듈 및 공간-스펙트럼 에이전트 및 시간 에이전트의 액션 및 이에 따른 보상에 관한 정보를 받고, 공간-스펙트럼 에이전트 및 시간 에이전트와 정책 및 가치를 교환하며 피드백을 수행하는 중앙 집중형 크리틱 네트워크 모듈을 포함할 수 있다.

[0009] 또한, 일 실시예에 따라서, 중앙 집중형 크리틱 네트워크 모듈은, $Q(s_t, \mathbf{a}_t) = \mathbb{E}_{(s_{t+1:T}, \mathbf{a}_{t+1:T})}[\mathcal{R}_t | s_t, \mathbf{a}_t]$ (여기서 S_t 는 현재 스텝에서 주어진 특징 전체 집합에 대한 상태, $\mathbf{a}_t$ 는 공간-스펙트럼 에이전트와 시간 에이전트가 수행한 액션을 합친 공동 액션, R_t 는 현재 상태에서 공동 액션을 수행하였을 때 받게 될 보상의 총합, $\mathbb{E}$ 는 확률 변수에 대한 평균 값을 의미한다.)의 공동 액션 가치 함수에 기반하여 피드백을 수행하는 것을 특징으로 할 수 있다.

[0010] 또한, 일 실시예에 따라서, 중앙 집중형 크리틱 네트워크 모듈은,

[0011]
$$A^i(s_t, \mathbf{a}_t) = Q(s_t, \mathbf{a}_t; \psi) - \sum_{\tilde{\mathbf{a}}_t^i} \pi_t^i(\tilde{\mathbf{a}}_t^i | o_t^i; \theta^i) Q(s_t, (\mathbf{a}_t^{-i}, \tilde{\mathbf{a}}_t^i; \psi)$$
 (여기서 ψ 는 업데이트가 진행되는 중앙집중형 크리틱의

파라미터, π_t^i 는 해당 에이전트 i 가 가지고 있는 정책, $\tilde{a}_t^i$ 는 해당 에이전트 i 가 선택할 수 있는 다른 종류의 액션, 0_t^i 는 에이전트 i 의 관찰, θ^i 는 에이전트 i 의 파라미터, a_t^{-i} 는 해당 에이전트 i 를 제외한 다른 에이전트의 모든 액션을 고정시키는 것을 의미한다.)의 가치 함수를 상기 공간-스펙트럼 에이전트 및 상기 시간 에이전트에 할당되도록 하는 것을 특징으로 할 수 있다.

[0012] 또한, 일 실시예에 따라서, 중앙 집중형 크리틱 네트워크 모듈은, 가치 함수를 최대화하는 방향으로 파라미터 θ^i 를 업데이트하는 것을 특징으로 할 수 있다.

[0013] 본 발명의 또 다른 실시예는 멀티에이전트 강화학습을 활용한 동작 상상 뇌파 분류 방법에 관한 것으로, 뇌파 데이터를 전처리하는 전처리 단계, 전처리된 뇌파 데이터의 복수의 도메인에서 의미 있는 특징을 전체 특징 집합으로 추출하는 단계, 복수의 에이전트를 활용하여 각 에이전트가 상기 복수의 도메인에 대해 관찰 및 액션을 수행하고 피드백 하는 멀티에이전트 강화학습을 통해 전체 특징 집합 중 동작 상상에 의미 있는 주요 특징을 추출하는 단계, 주요 특징을 통해 상기 뇌파 데이터로부터 사용자의 움직임 의도를 판단함으로써 동작 상상을 분류하는 단계를 포함할 수 있다.

[0014] 또한, 일 실시예에 따라서, 주요 특징을 추출하는 단계는, 공간-스펙트럼 에이전트를 통해 전체 특징 집합의 공간-스펙트럼 도메인 관찰을 수행하고, 정책을 수립하고, 액션을 결정 및 수행하는 공간-스펙트럼 특징 선택 단계, 시간 에이전트를 통해 전체 특징 집합의 시간 도메인 관찰을 수행하고, 정책을 수립하고, 액션을 결정 및 수행하는 시간 특징 선택 단계 및 공간-스펙트럼 에이전트 및 시간 에이전트의 액션 및 이에 따른 보상에 관한 정보를 받고, 공간-스펙트럼 에이전트 및 시간 에이전트와 정책 및 가치를 교환하며 피드백을 수행하는 중앙 집중형 크리틱 단계를 포함할 수 있다.

[0015] 또한, 일 실시예에 따라서, 중앙 집중형 크리틱 단계는, $Q(s_t, a_t) = \mathbb{E}_{(s_{t+1:T}, a_{t+1:T})}[\mathcal{R}_t | s_t, a_t]$ (여기서 S_t 는 현재 스텝에서 주어진 특징 전체 집합에 대한 상태, a_t 는 공간-스펙트럼 에이전트와 시간 에이전트가 수행한 액션을 합친 공동 액션, R_t 는 현재 상태에서 공동 액션을 수행하였을 때 받게 될 보상의 총합, $\mathbb{E}$ 는 확률 변수에 대한 평균 값을 의미한다.)의 공동 액션 가치 함수에 기반하여 피드백을 수행하는 것을 특징으로 할 수 있다.

[0016] 또한, 일 실시예에 따라서, 중앙 집중형 크리틱 단계는,

[0017] $A^i(s_t, a_t) = Q(s_t, a_t; \psi) - \sum_{\tilde{a}_t^i} \pi_t^i(\tilde{a}_t^i | o_t^i; \theta^i) Q(s_t, (a_t^{-i}, \tilde{a}_t^i); \psi)$ (여기서 ψ 는 업데이트가 진행되는 중앙집중형 크리틱의

파라미터, π_t^i 는 해당 에이전트 i 가 가지고 있는 정책, $\tilde{a}_t^i$ 는 해당 에이전트 i 가 선택할 수 있는 다른 종류의 액션, 0_t^i 는 에이전트 i 의 관찰, θ^i 는 에이전트 i 의 파라미터, a_t^{-i} 는 해당 에이전트 i 를 제외한 다른 에이전트의 모든 액션을 고정시키는 것을 의미한다.)의 가치 함수를 상기 공간-스펙트럼 에이전트 및 상기 시간 에이전트에 할당되도록 하는 것을 특징으로 할 수 있다.

[0018] 또한, 일 실시예에 따라서, 중앙 집중형 크리틱 단계는, 가치 함수를 최대화하는 방향으로 파라미터 θ^i 를 업데이트하는 것을 특징으로 할 수 있다.

발명의 효과

[0019] 본 발명의 동작 상상 뇌파 분류 장치 및 방법은, 특징 선택 과정에서 도메인적 지식 부족으로 인한 문제를 해결하는데 도움을 줄 수 있는 효과가 있다.

[0020] 또한, 본 발명의 동작 상상 뇌파 분류 장치 및 방법은, 강화학습을 활용하여 에이전트가 환경과 상호작용하며 보상을 최대화한 특징 조합을 스스로 학습함으로써, 특화된 도메인 지식에 크게 의존하지 않고도 보다 정확하고 유의미한 특징을 선택할 수 있는 효과가 있다.

[0021] 또한, 본 발명의 동작 상상 뇌파 분류 장치 및 방법은, 멀티에이전트 강화학습을 활용하여 다수의 에이전트가 협력하고 상호작용함으로써 더 나은 특징 조합을 찾아갈 수 있다.

[0022] 또한, 본 발명의 동작 상상 너과 분류 장치 및 방법은, 사용자의 동적 의도를 더욱 정확하게 예측하고 높은 수준의 제어와 상호작용을 실현할 수 있어, 다양한 응용 분야에서의 활용이 가능할 수 있다.

도면의 간단한 설명

[0023] 도 1은 일 실시예에 따른 동작 상상 너과 분류 장치의 블록도이다.

도 2는 일 실시예에 따른 학습부의 블록도이다.

도 3은 일 실시예에 따른 에이전트가 액션을 결정하는 과정에 대한 순서도이다.

도 4는 일 실시예에 따른 멀티에이전트 강화학습 과정에 대한 순서도이다.

도 5는 일 실시예에 따른 멀티에이전트 강화학습 과정에 대한 블록도이다

도 6은 일 실시예에 따른 각 방법들의 분류 정확도를 비교한 표이다.

도 7은 일 실시예에 따른 각 방법들의 분류 결과 지형도이다.

도 8은 일 실시예에 따른 동작 상상 너과 분류 방법의 순서도이다.

발명을 실시하기 위한 구체적인 내용

[0024] 본 발명의 이점 및 특징, 그리고 그것들을 달성하는 방법은 첨부되는 도면과 함께 후술되어 있는 실시예들을 참조하면 명확해질 것이다. 그러나 본 발명은 이하에서 개시되는 실시예들에 한정되는 것이 아니라 서로 다른 다양한 형태로 구현될 수 있으며, 단지 본 실시예들은 본 발명의 개시가 완전하도록 하고, 본 발명이 속하는 기술 분야에서 통상의 지식을 가진 자에게 발명의 범주를 완전하게 알려주기 위해 제공되는 것이며, 본 발명은 청구항의 범주에 의해 정의될 뿐이다.

[0025] 본 명세서에서 사용되는 용어에 대해 간략히 설명하고, 본 발명에 대해 구체적으로 설명하기로 한다.

[0026] 본 발명에서 사용되는 용어는 본 발명에서의 기능을 고려하면서 가능한 현재 널리 사용되는 일반적인 용어들을 선택하였으나, 이는 당 분야에 종사하는 기술자의 의도 또는 관례, 새로운 기술의 출현 등에 따라 달라질 수 있다. 또한, 특정한 경우는 출원인이 임의로 선정한 용어도 있으며, 이 경우 해당되는 발명의 설명 부분에서 상세히 그 의미를 기재할 것이다. 따라서 본 발명에서 사용되는 용어는 단순한 용어의 명칭이 아닌, 그 용어가 가지는 의미와 본 발명의 전반에 걸친 내용을 토대로 정의되어야 한다.

[0027] 명세서 전체에서 어떤 부분이 어떤 구성요소를 "포함"한다고 할 때, 이는 특별히 반대되는 기재가 없는 한 다른 구성요소를 제외하는 것이 아니라 다른 구성요소를 더 포함할 수 있음을 의미한다. 또한, 명세서에서 사용되는 "부", "모듈", "유닛" 등의 용어는 적어도 하나의 기능 또는 동작을 처리하는 단위를 의미하며, 소프트웨어, FPGA 또는 ASIC과 같은 하드웨어 구성요소, 또는 소프트웨어와 하드웨어의 결합으로 구현될 수 있다. 그렇지만 "부", "모듈", "유닛" 등의 용어가 소프트웨어 또는 하드웨어에 한정되는 의미는 아니다. "부", "모듈", "유닛" 등은 어드레싱할 수 있는 저장 매체에 있도록 구성될 수도 있고 하나 또는 그 이상의 프로세서들을 재생시키도록 구성될 수도 있다. 따라서, 일 예로서 "부", "모듈", "유닛" 등의 용어는 소프트웨어 구성요소들, 객체지향 소프트웨어 구성요소들, 클래스 구성요소들 및 태스크 구성요소들과 같은 구성요소들과, 프로세스들, 함수들, 속성들, 프로시저들, 서브루틴들, 프로그램 코드의 세그먼트들, 드라이버들, 펌웨어, 마이크로 코드, 회로, 데이터, 데이터베이스, 데이터 구조들, 테이블들, 어레이들 및 변수들을 포함한다.

[0028] 아래에서는 첨부한 도면을 참고하여 본 발명의 실시예에 대하여 본 발명이 속하는 기술 분야에서 통상의 지식을 가진 자가 용이하게 실시할 수 있도록 상세히 설명한다. 그리고 도면에서 본 발명을 명확하게 설명하기 위해서 설명과 관계없는 부분은 생략한다. "제1", "제2" 등과 같이 서수를 포함하는 용어는 다양한 구성 요소들을 설명하는데 사용될 수 있지만, 구성 요소들은 용어들에 의해 한정되지는 않는다. 용어들은 하나의 구성요소를 다른 구성요소로부터 구별하는 목적으로만 사용된다. 예를 들어, 본 발명의 권리 범위를 벗어나지 않으면서 제1 구성요소는 제2 구성요소로 명명될 수 있고, 유사하게 제2 구성요소도 제1 구성요소로 명명될 수 있다. "및/또는" 이라는 용어는 복수의 관련된 항목들의 조합 또는 복수의 관련된 항목들 중의 어느 하나의 항목을 포함한다.

[0029] 본 명세서에 설명된 다양한 시스템, 장치, 저장 매체, 모듈 및 기구는 전용 연산 디바이스, 또는 하나 이상의 전용 연산 디바이스의 하나 이상의 연산 칩에서 구현될 수도 있다.

- [0030] 이하, 본 발명의 멀티에이전트 강화학습을 활용한 동작 상상 뇌파 분류 장치 및 방법에 관하여 설명하도록 한다.
- [0031] 일 실시예에 의한 멀티에이전트 강화학습을 활용한 동작 상상 뇌파 분류 장치 및 방법은 뇌파 데이터를 입력받고 전처리할 수 있다. 또한, 동작 상상 뇌파 분류 장치 및 방법은 전처리된 뇌파 데이터의 복수의 도메인에서 의미 있는 특징을 전체 특징 집합으로 추출할 수 있다. 또한 동작 상상 뇌파 분류 장치 및 방법은 복수의 에이전트를 활용하여 각 에이전트가 전체 특징 집합 내 복수의 도메인에 대해 관찰 및 액션을 수행하고 피드백 하는 멀티에이전트 강화학습을 통해 전체 특징 집합 중 동작 상상에 의미 있는 주요 특징을 추출할 수 있다. 또한, 동작 상상 뇌파 분류 장치 및 방법은 주요 특징을 통해 뇌파 데이터로부터 사용자의 움직임 의도를 판단하여 동작 상상을 분류할 수 있다.
- [0032] 이하 도면을 참조하여 본 발명의 일 실시예에 의한 동작 상상 뇌파 분류 장치 및 방법에 관하여 설명하도록 한다.
- [0033] 도 1은 본 발명의 동작 상상 뇌파 분류 장치(1)의 블록도이다.
- [0034] 도 1의 예시에 의하면, 동작 상상 뇌파 분류 장치(1)는 입력부(10), 저장부(20), 출력부(30) 및 프로세서(40)를 포함할 수 있다. 이는 일 실시예에 의한 구성들을 예시한 것이며 필요에 따라 이외의 다른 구성을 더 포함할 수 있다. 동작 상상 뇌파 분류 장치(1)의 각 구성은 케이블, 회선 또는 무선통신 네트워크를 통해 일방으로 또는 쌍방으로 데이터를 전달할 수 있도록 마련될 수 있다.
- [0035] 입력부(10)는 본 발명의 동작 상상 뇌파 분류 장치(1)의 동작에 필요한 명령, 데이터 및 프로그램 중 적어도 하나를 입력 받고, 이를 프로세서(40) 및/혹은 저장부(20)로 전달할 수 있다. 예를 들어, 입력부(10)는 뇌파 데이터를 입력 받을 수 있다.
- [0036] 뇌파 데이터는 뇌에서 발생하는 전기 활동을 측정하여 획득한 데이터로서, 주로 EEG(electroencephalography) 기술을 활용하여 수집한 EEG(electroencephalogram) 뇌파 데이터를 의미하는 것일 수 있다. 뇌파 데이터는 주로 일정 시간 동안 뇌에서 발생하는 전기 활동을 시각적으로 나타낸 그래프로 표현될 수 있으며, 그래프를 통해 뇌의 다양한 영역에서 발생하는 뇌파 주파수 범위(이하, '스펙트럼'과 혼용될 수 있음)의 변화를 관찰할 수 있다. 즉, 뇌파 데이터는 뇌의 신경세포들이 서로 통신하고 정보를 처리하는 과정에서 발생하는 전기적 신호를 수집한 것으로, 전기적 신호의 시간적, 공간적 및 주파수적 정보를 포함하고 있을 수 있다.
- [0037] 저장부(20)는 명령, 데이터 또는 프로그램 등을 일시적 또는 비일시적으로 저장할 수 있다. 이러한 명령, 데이터 또는 프로그램 등은 입력부(10)를 통해 입력된 것일 수도 있다. 예를 들어, 뇌파 데이터 및 강화학습에 관한 데이터를 포함할 수 있다. 또한, 저장부(20)에 저장되는 데이터는 프로세서(40)의 처리 과정에서 생성된 것, 일례로 후술될 전처리된 뇌파 데이터나 전체 특징 집합에 관한 데이터 등을 포함할 수도 있다. 저장부(20)는 프로세서(40)의 호출에 따라 저장된 데이터 등을 프로세서(40)에 제공하는 것도 가능하다. 저장부(20)는, 주기억장치 및 보조기억장치 중 적어도 하나를 포함할 수 있으며, 이들은 반도체나 자기디스크 등을 이용해 구현될 수 있다.
- [0038] 또한, 저장부(20)는 인공지능 모델을 저장할 수 있다. 본 발명에서 사용될 수 있는 인공지능 모델은 적어도 하나의 태스크를 학습하는 기계학습 모델로, 프로세서(40)에 의해 실행되는 컴퓨터 프로그램 등으로 구현될 수 있다. 여기서 인공지능 모델이 학습하는 태스크란, 기계학습을 통해 해결하고자 하는 과제 또는 기계학습을 통해 수행하고자 하는 작업을 지칭할 수 있다. 인공지능 모델은 컴퓨팅 장치에서 실행되는 컴퓨터 프로그램으로 구현될 수 있고, 네트워크를 통해 다운로드 될 수도 있다.
- [0039] 일 실시예에 의하면, 본 발명의 인공지능 모델은 강화학습에 활용될 수 있는 에이전트를 포함할 수 있다. 에이전트는 환경과 상호작용하며, 현재의 상태(state)를 관찰하고, 가능한 액션(action) 중에서 가장 적절한 액션을 선택하여 수행할 수 있다. 또한, 에이전트는 환경으로부터 얻는 보상(reward)과 벌점에 따라 액션의 성과를 평가하고, 이를 기반으로 정책(policy)을 개선하여 더 나은 액션을 수행할 수 있도록 할 수 있다. 즉, 본 발명의 인공지능 모델은 환경과 상호작용하여 보상과 벌점을 통해 스스로 학습하는 강화학습 에이전트를 포함할 수 있으며, 에이전트는 현재 상태를 관찰하고 적절한 액션을 선택하여 수행하며, 보상에 따라 정책을 개선하는 형태로 학습할 수 있다.
- [0040] 출력부(30)는 외부로 데이터나 프로그램 등을 출력할 수 있게 마련된다. 예를 들어, 출력부(30)는 후술할 주요 특징이나, 동작 상상 뇌파 분류 결과 등을 출력하여 사용자에게 제공하거나 다른 정보처리장치로 전송할 수도

있다. 출력부(30)는, 예를 들어, 디스플레이, 프린터 장치, 스피커 장치, 영상 출력 단자, 데이터 입출력 단자 또는 통신 모듈 등을 포함할 수 있으나, 이에 한정되는 것은 아니다.

[0041] 프로세서(40)는 입력부(10) 및 저장부(20)에서 뇌파 데이터 등을 전달받을 수 있고, 전달받은 뇌파 데이터를 전처리할 수 있다. 또한, 프로세서(40)는 전처리된 뇌파 데이터로부터 복수의 도메인에서 의미 있는 특징을 전체 특징 집합으로 추출할 수 있다. 또한, 프로세서(40)는 복수의 에이전트를 활용하여 복수의 도메인에 대해 관찰 및 액션을 수행하고 피드백 하는 멀티에이전트 강화학습을 통해 전체 특징 집합 중 동작 상상에 의미 있는 주요 특징을 추출할 수 있다. 또한, 프로세서(40)는 주요 특징을 통해 뇌파 데이터로부터 사용자의 움직임 의도를 판단하여 동작 상상을 분류할 수 있다.

[0042] 프로세서(40)는 실시예에 따라 전술한 과정을 모두 수행할 수 있으며, 이 중 일부만을 수행할 수도 있다. 이들 동작의 수행을 위해 저장부(20) 등에 미리 저장된 프로그램이 이용될 수 있다. 여기서 프로그램은 설계자 등에 의해 직접 작성된 것일 수 있고, 전자소프트웨어 유통망을 통해 획득 또는 갱신된 것을 수 있다.

[0043] 실시예에 의하면, 프로세서(40)는, 중앙 처리 장치(CPU: Central Processing Unit), 그래픽 처리 장치(GPU: Graphic Processing Unit), 마이크로 컨트롤러 유닛(MCU: Micro Controller Unit), 애플리케이션 프로세서(AP: Application Processor), 전자 제어 유닛(ECU: Electronic Controlling Unit), 마이크로컴(Micom: Micro Processor) 또는 이외 각종 연산 및 제어 처리를 수행할 수 있는 적어도 하나의 전자 장치 등을 하나 이상 포함할 수 있다. 이들 처리 또는 제어 장치는, 예를 들어, 하나 또는 둘 이상의 반도체 칩, 회로 또는 관련 부품 등을 단독으로 이용하거나 조합하여 구현된 것일 수도 있다.

[0044] 일 실시예에 의하면, 프로세서(40)는 전처리부(400), 특징 추출부(420), 학습부(440) 및 분류부(460)를 포함할 수 있다. 여기서, 각 구성은 물리적 및 논리적으로 구분되는 것일 수 있으며, 이들은 하나 또는 둘 이상의 처리 장치(중앙처리장치 및 그래픽처리장치 등)를 이용하여 구현될 수 있다. 실시예에 따라 프로세서(40)의 각 구성 중 일부는 생략될 수 있으며, 필요에 따라서는 전술한 구성 이외의 구성이 더 포함될 수 있다.

[0045] 전처리부(400)는 입력부(10)로부터 뇌파 데이터를 전달받고, 이를 전처리할 수 있다. EEG 뇌파 데이터는 사람들마다 개별차이가 심하고, 여러가지 노이즈가 포함되어 있을 수 있기 때문에 이를 분석에 활용하기 전에 전처리가 필요할 수 있다.

[0046] 전처리부(400)는 여러가지 전처리 기법을 이용하여 뇌파 데이터를 전처리할 수 있다. 예를 들어, 전처리부(400)는 잡음 제거, 베이스라인 보정, 아티팩트 제거, 시간 정렬과 같은 전처리 기법을 활용할 수 있다. 여기서 잡음 제거는 원하는 주파수 대역의 뇌파 신호를 보존하고 그 외의 주파수를 제거하는 대역 통과 필터링(band-pass filtering)과 같은 방법을 통해 수행될 수 있다. 전처리부(400)에서 활용될 수 있는 전처리 기법은 앞선 예시들에 한정되는 것은 아니며, 일반적으로 EEG 뇌파 데이터의 전처리에 사용되는 방법은 제한없이 활용될 수 있다.

[0047] 특징 추출부(420)는 전처리부(400)로부터 전처리된 뇌파 데이터를 전달받고, 이에 기초하여 복수의 도메인에서 의미 있는 특징을 전체 특징 집합으로 추출할 수 있다. 여기서 의미 있는 특징은, 뇌파 데이터에 포함된 정보에서 기계 학습에 활용될 수 있을 정도로 특정되어 의미를 가질 수 있는 특징들을 의미할 수 있다.

[0048] 전처리된 뇌파 데이터는 전술한 바와 같이 복수의 도메인과 관련된 정보를 포함하고 있을 수 있다. 일 실시예에 의하면, 뇌파 데이터는 시간 도메인, 공간 도메인, 주파수 도메인과 관련된 정보를 포함하고 있을 수 있다. 특징 추출부(420)는 예시한 바와 같은 복수의 도메인과 관련된 정보에서 의미 있는 특징들을 모두 추출할 수 있으며, 이들 집합을 전체 특징 집합으로 정의할 수 있다. 일 실시예에 의하면, 특징 추출부(420)는 시간 도메인 및 공간-스펙트럼 도메인(공간 도메인 및 주파수 도메인)에서 의미 있는 특징을 모두 추출하여 전체 특징 집합을 생성할 수 있으며, 전체 특징 집합은 텐서 형태로 생성될 수 있다. 일 실시예에 의하면, 전체 특징 집합은 텐서 형태로, 가로축은 시간 도메인을 나타낼 수 있고, 세로축은 공간-스펙트럼 도메인을 나타내도록 생성될 수 있다. 또한, 전체 특징 집합은 이전 특징 및 현재 특징을 모두 포함하는 형태로 생성될 수 있다.

[0049] 도 2 내지 도 5를 참조하여 학습부(440)에 관하여 설명하도록 한다.

[0050] 학습부(440)에서는 강화학습을 활용하여 주요 특징을 추출할 수 있다. 여기서 주요 특징은, 뇌파 데이터에서 추출한 전체 특징 중 동작 상상과 관련하여 의미를 갖는 중요한 특징들을 의미할 수 있다. 예를 들어, 추출한 전체 특징 중에서 '오른손을 움직인다'의 동작 의도와 관련된 특징 등을 의미할 수 있다.

[0051] 학습부(440)는 강화학습을 활용할 수 있다. 강화학습은 기계학습의 한 분야로 에이전트 주체가 환경과 상호작용하며 정해진 목표를 달성하기 위한 액션을 스스로 학습하는 알고리즘이다. 에이전트는 상태를 관찰하여 액션을

결정 및 수행하여 리워드를 산출하고, 이를 최대화하도록 업데이트하는 과정을 통해 최적의 정책을 학습할 수 있다.

- [0052] 또한, 학습부(440)는 강화학습 중 멀티에이전트 강화학습을 활용할 수 있다. 멀티에이전트 강화학습은 복수의 에이전트를 활용하며, 에이전트간 상호작용을 통해 전술한 과정을 거쳐 최적의 정책을 학습하는 알고리즘일 수 있다.
- [0053] 학습부(440)는 특징 추출부(420)로부터 전체 특징 집합을 전달받고, 복수의 에이전트를 활용하여 각 에이전트가 전체 특징 집합 내 서로 다른 도메인에 대하여 관찰 및 액션을 수행하도록 할 수 있다. 또한, 학습부(440)는 멀티에이전트 강화학습을 통해 전체 특징 집합 중 동작 상상에 의미 있는 주요 특징을 추출할 수 있다.
- [0054] 일 실시예에 의하면, 복수의 에이전트는 공간-스펙트럼 에이전트 및 시간 에이전트를 포함하여, 뇌파 데이터 내 공간-스펙트럼 도메인 및 시간 도메인을 관찰하고, 관찰에 대응하는 액션을 수행할 수 있다.
- [0055] 도 2는 일 실시예에 의한 학습부(440)의 구성을 나타내는 블록도이다.
- [0056] 일 실시예에 의한 학습부(440)는 상태 모듈(441), 공간-스펙트럼 특징 모듈(445), 시간 추출 모듈, 중앙 집중형 크리티크 네트워크 모듈(449) 및 리워드 모듈(443)을 포함할 수 있다.
- [0057] 상태 모듈(441) 및 리워드 모듈(443)은 에이전트와 상호작용하는 환경에 관한 정보를 제어할 수 있다.
- [0058] 상태 모듈(441)은 전달받은 전체 특징 집합을 현재 상태로 정의할 수 있으며, 상태에 관한 정보를 생성, 업데이트, 저장 및 전달할 수 있다.
- [0059] 도 3을 참조하여 공간-스펙트럼 특징 모듈(445) 및 시간 특징 모듈(447)에 관하여 설명하도록 한다.
- [0060] 각 특징 모듈에서 에이전트는 현재 상태에서 각 도메인을 관찰하여 이에 대응하는 특징을 추출할 수 있다. 또한, 각 특징 모듈에서 에이전트는 이를 기반으로 하여 액션을 결정 및 수행할 수 있다.
- [0061] 도 3은 에이전트가 액션을 결정하는 과정을 나타낸 순서도이다.
- [0062] 공간-스펙트럼 특징 모듈(445)에서, 공간-스펙트럼 에이전트는 현재 상태의 공간-스펙트럼 도메인을 관찰하고, 이에 대한 액션을 수행할 수 있다.
- [0063] 보다 상세하게는, 공간-스펙트럼 특징 모듈(445)에서 공간-스펙트럼 에이전트는 정의된 현재 상태의 공간-스펙트럼 도메인에서 이전 특징과 현재 특징을 모두 관찰할 수 있다(S310).
- [0064] 또한, 공간-스펙트럼 특징 모듈(445)에서 공간-스펙트럼 에이전트는 정책을 수립할 수 있다(S320). 구체적으로, 공간-스펙트럼 에이전트는 이전에 관찰한 현재 특징이 동적 상상 분류에 도움이 되는지 여부를 고려하여 정책을 스스로 수립하도록 학습할 수 있다. 정책은 후술할 중앙 집중형 크리티크 네트워크 모듈(449)에서 제공하는 공동 액션 가치 함수에 기반하여 업데이트 될 수 있다.
- [0065] 또한, 공간-스펙트럼 특징 모듈(445)에서 공간-스펙트럼 에이전트는 특징 평가를 수행할 수 있다(S330). 공간-스펙트럼 에이전트는 수립한 정책과 후술할 중앙 집중형 크리티크 네트워크 모듈(449)에서 제공하는 액션 가치에 근거하여 현재 특징이 보상을 최대화하는지를 판단함으로써 특징의 유용성을 판단할 수 있다.
- [0066] 또한, 공간-스펙트럼 특징 모듈(445)에서 공간-스펙트럼 에이전트는 액션을 결정 및 수행할 수 있다(S340). 공간-스펙트럼 에이전트는 전술한 특징 평가의 결과를 기반으로 현재 특징을 채택하거나 채택하지 않을 수 있다.
- [0067] 시간 특징 모듈(447)에서, 시간 에이전트는 현재 상태의 시간 도메인을 관찰하고, 이에 대한 액션을 수행할 수 있다.
- [0068] 보다 상세하게는, 시간 특징 모듈(447)에서 시간 에이전트는 정의된 현재 상태의 시간 도메인에서 이전 특징과 현재 특징을 모두 관찰할 수 있다(S310).
- [0069] 또한, 시간 특징 모듈(447)에서 시간 에이전트는 정책을 수립할 수 있다(S320). 구체적으로, 시간 에이전트는 이전에 관찰한 현재 특징이 동적 상상 분류에 도움이 되는지 여부를 고려하여 정책을 스스로 수립하도록 학습할 수 있다. 정책은 후술할 중앙 집중형 크리티크 네트워크 모듈(449)에서 제공하는 공동 액션 가치 함수에 기반하여 업데이트 될 수 있다.
- [0070] 또한, 시간 특징 모듈(447)에서 시간 에이전트는 특징 평가를 수행할 수 있다(S330). 시간 에이전트는 수립한 정책과 후술할 중앙 집중형 크리티크 네트워크 모듈(449)에서 제공하는 액션 가치에 근거하여 현재 특징이 보상을

최대화하는지를 판단함으로써 특징의 유용성을 판단할 수 있다.

- [0071] 또한, 시간 특징 모듈(447)에서 시간 에이전트는 액션을 결정 및 수행할 수 있다(S340). 이에 따라 시간 에이전트는 전술한 특징 평가의 결과를 기반으로 현재 특징을 채택하거나 채택하지 않을 수 있다.
- [0072] 이상으로, 공간-스펙트럼 특징 모듈(445)의 공간-스펙트럼 에이전트 및 시간 특징 모듈(447)의 시간 에이전트의 예를 들어 각 에이전트가 액션을 결정하고 수행하는 과정에 관하여 설명하였으나, 이에 한정되는 것은 아니며, 본 발명의 장치에서 활용하고자 하는 도메인에 따라 명칭 및 기능은 변경될 수 있다.
- [0073] 리워드 모듈(443)은 보상에 관한 정보를 생성, 업데이트, 저장 및 전달할 수 있다.
- [0074] 리워드 모듈(443)은 각 에이전트의 액션에 대한 보상을 산출하고, 이를 후술할 중앙 집중형 크리티크 네트워크 모듈(449)에 제공할 수 있다. 즉, 일 실시예에 의하면, 리워드 모듈(443)은 공간-스펙트럼 에이전트의 액션 및 시간 에이전트의 액션을 모두 고려된 공동 액션에 따라 보상을 산출하고, 피드백을 위하여 이를 중앙 집중형 크리티크 네트워크 모듈(449)에 제공할 수 있다.
- [0075] 중앙 집중형 크리티크 네트워크 모듈(449)은 각 특징 모듈과의 상호작용을 통해 각 에이전트가 동작 상상 분류에 더 적합한 특징을 채택할 수 있도록 할 수 있다. 중앙 집중형 크리티크 네트워크 모듈(449)은 멀티에이전트 강화 학습을 위해 필요한 구성으로, 복수의 에이전트와 상호작용하며 모든 도메인에 대한 정보를 고려하도록 하여 최적의 정책을 수립하도록 피드백을 제공할 수 있다.
- [0076] 중앙 집중형 크리티크 네트워크 모듈(449)은 현재 상태에 대한 모든 정보를 관찰할 수 있다. 일 실시예에 의하면, 중앙 집중형 크리티크 네트워크 모듈(449)은 현재 상태의 공간-스펙트럼 도메인 및 시간 도메인을 포함한 모든 도메인의 정보를 관찰할 수 있다. 이는 중앙 집중형 크리티크 네트워크의 경우, 각 에이전트와 달리 전반적인 상황을 고려해야하기 때문이다.
- [0077] 중앙 집중형 크리티크 네트워크 모듈(449)은 각 특징 모듈과 상호작용하며 피드백을 제공할 수 있다. 또한, 중앙 집중형 크리티크 네트워크 모듈(449)은 각 에이전트가 수행하는 액션에 대한 정책을 비평하고 가치를 피드백으로 제공할 수 있다. 즉, 일 실시예에 의하면, 중앙 집중형 크리티크 네트워크 모듈(449)은 공간-스펙트럼 특징 모듈(445)의 공간-스펙트럼 에이전트와 상호작용을 통해 정책을 비평하고 가치를 피드백으로 제공할 수 있다. 또한, 마찬가지로 중앙 집중형 크리티크 네트워크 모듈(449)은 시간 특징 모듈(447)의 시간 에이전트와 상호작용을 통해 정책을 비평하고 가치를 피드백으로 제공할 수 있다.
- [0078] 일 실시예에 의하면, 중앙 집중형 크리티크 네트워크는 각 특징 모듈에 공동 액션 가치 함수를 제공할 수 있다.
- [0079] 일 실시예에 의하면 공동 액션 가치 함수는 다음 수학식 1과 같이 정의될 수 있다.

수학식 1

$$Q(s_t, \mathbf{a}_t) = \mathbb{E}_{(s_{t+1:T}, \mathbf{a}_{t+1:T})}[\mathcal{R}_t | s_t, \mathbf{a}_t]$$

- [0080]
- [0081] 여기서 S_t 는 현재 스텝에서 주어진 특징 전체 집합에 대한 상태를 의미하며, $\mathbf{a}_t$ 는 공간-스펙트럼 에이전트와 시간 에이전트가 수행한 액션을 합친 공동 액션을 의미한다. R_t 는 현재 상태에서 공동 액션을 수행하였을 때 받게 될 보상의 총합을 의미한다. $\mathbb{E}$ 는 확률론에서 사용되는 기댓값으로 확률 변수에 대한 평균적인 결과 또는 평균 값을 의미한다.
- [0082] 이와 같이 중앙 집중형 크리티크 네트워크 모듈(449)은 공동 액션 가치 함수를 통해 상태와 공동 액션을 입력으로 받아 해당 상태에서 해당 공동 액션을 선택했을 때 얼마나 좋은 결과를 기대할 수 있는지를 평가할 수 있다.
- [0083] 이렇게 정의된 공동 액션 가치 함수를 사용하여 중앙 집중형 크리티크 네트워크 모듈(449)은 다음 수학식 2와 같이 업데이트 될 수 있다.

수학식 2

$$\mathcal{L}_t(\psi) = \mathbb{E}_{(s_t, \mathbf{a}_t, r_t, s_{t+1}, \mathbf{a}_{t+1}) \sim B} [(Q(s_t, \mathbf{a}_t; \psi) - y)^2],$$

$$y = r_t + \gamma(Q(s_{t+1}, \mathbf{a}_{t+1}; \psi)),$$

[0084]

[0085]

여기서 ψ 는 업데이트가 진행되는 중앙집중형 크리틱의 파라미터를 의미한다. r_t 는 현재 스텝에서 받는 리워드를 의미하며 γ 는 MDP 모델링에서 할인 요소를 나타낸다. B 는 버퍼로써 데이터를 임시로 저장하거나 처리하는 데 사용되는 메모리를 의미한다. MDP 모델링에 관한 구체적인 설명은 본 발명에서는 생략하도록 한다.

[0086]

앞서 정의한 공동 액션 가치 함수를 바탕으로 각각의 에이전트에 할당되는 가치 함수는 다음과 같이 수학식 3으로 표현 가능하다.

수학식 3

$$A^i(s_t, \mathbf{a}_t) = Q(s_t, \mathbf{a}_t; \psi) - \sum_{\tilde{\mathbf{a}}_t^i} \pi_t^i(\tilde{\mathbf{a}}_t^i | o_t^i; \theta^i) Q(s_t, (\mathbf{a}_t^{-i}, \tilde{\mathbf{a}}_t^i; \psi)$$

[0087]

[0088]

이 수식에서 $\tilde{\mathbf{a}}_t^i$ 는 해당 에이전트 i 가 선택할 수 있는 다른 종류의 액션을 의미하며 $\mathbf{a}_t^{-i}$ 는 해당 에이전트 i 를 제외한 다른 에이전트의 모든 액션을 고정시키는 것을 의미한다. 추가로 π_t^i 는 해당 에이전트 i 가 가지고 있는 정책, o_t^i 는 에이전트 i 의 관찰, θ^i 는 에이전트 i 의 파라미터를 나타낸다. 이 정책을 기반으로 해당 에이전트는 액션을 선택하게 된다. 각각의 에이전트는 앞서 정의한 가치 함수를 최대화하는 방향으로 θ^i 를 업데이트한다. 이를 위한 수식은 다음 수학식 4와 같이 표현할 수 있다.

수학식 4

$$\mathcal{L}_t(\theta^i) = -\mathbb{E}_{\pi^i} [\nabla_{\theta^i} \log \pi^i(a_t^i | o_t^i; \theta^i) A^i(s_t, \mathbf{a}_t)]$$

[0089]

[0090]

여기서 π^i 는 에이전트가 자체적으로 수립한 정책을 나타내며 ∇ 는 벡터 또는 텐서 함수에 대한 미분 연산을 의미한다.

[0091]

이 업데이트는 반복적인 학습 단계를 통해 이루어질 수 있다. 각 에이전트는 상태 모듈(441), 리워드 모듈(443) 및 중앙 집중형 크리틱 네트워크 모듈(449)과 상호작용 하면서 최적의 정책을 수립하고, 최적의 행동을 수행하도록 학습할 수 있다. 중앙 집중형 크리틱 네트워크 모듈(449)은 다양한 에이전트로부터 수집된 정보를 활용하여 가치 함수를 업데이트하고, 이를 통해 에이전트는 더 효과적이고 협력적인 액션을 수행할 수 있다. 즉, 멀티 에이전트 강화학습은 분산된 에이전트들 간 협력과 정보 교환을 통해 전체 시스템의 성능을 향상시키며, 각 에이전트의 개별적인 액션 역시 최적화할 수 있다.

[0092]

도 4 및 도 5를 참조하여 학습부(440)에서 멀티에이전트 강화학습을 통해 주요 특징을 추출하는 과정에 관하여 설명하도록 한다.

[0093]

도 4는 학습부(440)에서 멀티에이전트 강화학습을 수행하는 과정의 순서도이다.

[0094]

도 4에 따르면, 멀티에이전트 강화학습 과정은 초기화 단계(S410)를 포함할 수 있다. 멀티에이전트 강화학습 과정은 학습을 시작하기전 각 에이전트의 상태를 초기화할 수 있다.

[0095]

또한, 멀티에이전트 강화학습 과정은 액션 선택 단계(S420)를 포함할 수 있다. 멀티에이전트 강화학습 과정에서, 각 에이전트는 현재의 상태를 기반으로 특징 선택 액션을 결정할 수 있다.

[0096]

또한, 멀티에이전트 강화학습 과정은 액션 수행 단계(S430)를 포함할 수 있다. 멀티에이전트 강화학습 과정은,

선택된 특징을 환경에 적용하여 다음 상태로 전이할 수 있다.

- [0097] 또한, 멀티에이전트 강화학습 과정은 보상 획득 단계(S440)를 포함할 수 있다. 멀티에이전트 강화학습 과정에서 각 에이전트는 리워드 모듈(443)을 통해 다음 상태로부터 받는 보상을 획득할 수 있다.
- [0098] 또한, 멀티에이전트 강화학습 과정은 정보 전달 단계(S450)를 포함할 수 있다. 멀티에이전트 강화학습 과정에서, 각 에이전트 및 환경은 중앙 집중형 크리틱 네트워크 모듈(449)에 현재 상태, 액션, 리워드, 다음 상태에 관한 정보를 전달할 수 있다.
- [0099] 또한, 멀티에이전트 강화학습 과정은 중앙 집중형 크리틱 단계(S460)를 포함할 수 있다. 멀티에이전트 강화학습 과정에서 중앙 집중형 크리틱 네트워크 모듈(449)은 획득한 정보를 기반으로 공동 액션 가치 함수를 업데이트하고, 각 에이전트의 액션 가치를 계산할 수 있다.
- [0100] 또한, 멀티에이전트 강화학습 과정은 정책 업데이트 단계(S470)를 포함할 수 있다. 멀티에이전트 강화학습 과정에서 각 에이전트들은 중앙 집중형 크리틱 네트워크 모듈(449)과 정보를 교환함으로써 최적의 액션을 선택하도록 각자의 정책을 스스로 업데이트할 수 있다.
- [0101] 또한, 멀티에이전트 강화학습 과정은 종료 여부 확인 단계(S480)를 포함할 수 있다. 멀티에이전트 강화학습 과정은, 종료 조건을 만족하는지 확인할 수 있다. 여기서 종료 조건은 예를 들어, 최대 에피소드 수 달성, 분류 목표 달성 등이 포함될 수 있다.
- [0102] 또한, 멀티에이전트 강화학습 과정은 종료 혹은 반복 단계(S490)를 포함할 수 있다. 멀티에이전트 강화학습 과정은, 앞선 종료 조건이 만족되지 않아 과정이 종료되지 않은 경우, 위의 액션 선택 단계로 돌아가 과정을 반복하고, 조건을 만족하면 과정을 종료할 수 있다.
- [0103] 전술한 과정을 거쳐 학습부(440)는 동작 상상 뇌파 분류에 의미 있는 주요 특징을 추출할 수 있다.
- [0104] 도 5는 앞선 멀티 에이전트 강화학습 과정을 간략하게 도시한 블록도이다.
- [0105] 도 5의 예시에 따르면, 상태 모듈(441)은 전달받은 전체 특징 집합을 현재 상태로 정의할 수 있다.
- [0106] 전체 특징 집합의 정의 이후, 각 에이전트는 상태의 일 도메인에서 관찰을 수행할 수 있다. 일 실시예에 의하면, 공간-스펙트럼 에이전트는 상태의 공간-스펙트럼 도메인에서 관찰을 수행할 수 있다. 또한, 마찬가지로 시간 에이전트는 상태의 시간 도메인에서 관찰을 수행할 수 있다.
- [0107] 에이전트는 관찰을 수행한 후, 이를 기초로 액션을 선택 및 수행할 수 있다. 각 에이전트는 관찰한 현재 상태의 일 도메인에 기초하여 특징 선택 액션을 결정하고, 결정된 액션을 수행할 수 있다.
- [0108] 일 실시예에 의하면, 공간-스펙트럼 에이전트는 현재 상태의 공간-스펙트럼 도메인에서 수행한 관찰 결과에 기초하여 특징 선택 액션을 결정하고, 결정된 액션을 수행할 수 있다. 또한, 마찬가지로 시간 에이전트는 현재 상태의 시간 스펙트럼 도메인에서 수행한 관찰 결과에 기초하여 특징 선택 액션을 결정하고, 결정된 액션을 수행할 수 있다.
- [0109] 각 에이전트에서 수행한 액션을 모두 고려한 공동 액션에 기초하여 리워드를 산출하고 이를 중앙 집중형 크리틱 네트워크 모듈(449)에 전달할 수 있다. 중앙 집중형 크리틱 네트워크 모듈(449)은 각 에이전트와 상호작용을 통해 얻은 정책 정보 및 상태를 관찰하여 획득한 현재 상태 정보 등을 이용하여 각 에이전트에 가치를 피드백으로 제공할 수 있다. 이에 따라 각 에이전트는 정책을 업데이트하고 최적화된 액션을 결정 및 수행할 수 있다. 이를 통해 학습부(440)는 전체 특징 집합 중 동작 상상 뇌파 분류에 의미 있는 주요 특징을 추출할 수 있다.
- [0110] 다시 도 1을 참조하여 분류부(460)에 관하여 설명하면, 분류부(460)는 학습부(440)에서 추출된 주요 특징을 전달받을 수 있다. 또한, 분류부(460)는 전달받은 주요 특징에 기초하여 사용자의 움직임 의도를 판단할 수 있다. 분류부(460)는 이를 기초로 사용자의 동작 상상을 분류할 수 있다.
- [0111] 즉, 분류부(460)는 동작 상상 뇌파 분류를 위해 입력으로 주어진 사용자의 뇌파 데이터를 학습부(440)에서 전달받은 주요 특징을 기반으로 분류함으로써, 사용자의 각각의 동작 의도에 대한 예측 결과를 분류할 수 있다. 예를 들어 분류부(460)는 '손을 움직인다'라는 동작 의도에 대한 주요 특징을 전달받고, 입력받은 사용자의 뇌파 데이터로부터 이와 유사한 특징이 산출되는 경우, 이를 기초로 사용자에게 손을 움직이는 동작 의도가 있는 것으로 판단할 수 있다.
- [0112] 도 6 및 도 7을 참조하여 본 발명의 일 실시예에 의한 동작 상상 뇌파 분류 장치(1)의 성능 테스트 결과에 관하

여 설명하도록 한다.

- [0113] 도 6은 기존의 동작 상상 예측 장치 및 본 발명의 동작 상상 뇌파 분류 장치(1)의 분류 정확도를 비교한 표이다.
- [0114] 도 6은 단순한 특징 추출기를 사용한 경우, 이에 더해 단일 에이전트를 사용한 경우, 이에 더해 복수의 단일 에이전트를 사용한 경우 및 이에 더해 멀티에이전트 강화학습 기법을 활용한 본 발명의 경우를 데이터셋 1 및 데이터셋 2를 이용하여 정확도를 테스트한 결과이다. 구체적으로, SAT는 시간 도메인에서 특징을 선택하는 에이전트이며, SAS는 공간-스펙트럼 도메인에서 특징을 선택하는 에이전트이다. 또한, 복수의 단일 에이전트를 사용하는 경우는 각 에이전트간 상호작용이 고려되지 않았다. 데이터 셋 1은 BCI Competition IV 2a이며 데이터 셋 2는 KU-MI이다.
- [0115] 도 6에 따르면, 특징 추출기를 사용한 경우 각 데이터 셋에 대한 분류 정확도는 69.59(±0.61) 및 64.99(±0.20)로 측정 되었다.
- [0116] 여기에 더해 SAT 단일 에이전트를 사용한 경우 각 데이터 셋에 대한 분류 정확도는 70.51(±0.37) 및 66.00(±0.23)로 측정되었으며, SAS 단일 에이전트를 더 사용한 경우 각 데이터 셋에 대한 분류 정확도는 70.80(±0.95) 및 66.04(±0.44)로 측정되었다.
- [0117] 여기에 더해 복수의 단일 에이전트 SAS 및 SAT를 사용한 경우, 각 데이터 셋에 대한 분류 정확도는 70.69(±0.60) 및 66.02(±0.30) 그리고 70.79(±0.40) 및 66.04(±0.43)로 측정되었다.
- [0118] 본 발명의 멀티에이전트 강화학습을 활용한 동작 상상 뇌파 분류 장치(1)의 경우, 각 데이터셋에 대한 분류 정확도는 72.13(±0.62) 및 66.74(±0.31)로 측정되었다.
- [0119] 이를 통해 본 발명의 멀티에이전트 강화학습을 활용한 동작 상상 뇌파 분류 장치(1)는 기존의 방법들에 비하여 분류 정확도가 향상된 것을 확인할 수 있다.
- [0120] 도 7은 기존의 동작 상상 예측 장치 및 본 발명의 동작 상상 뇌파 분류 장치(1)의 분류 결과 지형도이다.
- [0121] 도 7의 그래프를 비교하여 보았을 때, 본 발명의 동작 상상 뇌파 분류 장치(1)의 경우, 다른 방법들에 비하여 더 정확한 분류를 제공할 수 있음을 확인할 수 있다.
- [0122] 도 8을 참조하여 본 발명의 동작 상상 뇌파 분류 방법에 관하여 설명하도록 한다.
- [0123] 도 8은 본 발명의 동작 상상 뇌파 분류 방법의 순서도이다.
- [0124] 본 발명의 동작 상상 뇌파 분류 방법은 멀티에이전트 강화학습을 활용한 동작 상상 뇌파 분류 방법에 관한 것으로, 뇌파 데이터를 전처리하는 전처리 단계(S810), 전처리된 뇌파 데이터의 복수의 도메인에서 의미 있는 특징을 전체 특징 집합으로 추출하는 단계(S820), 복수의 에이전트를 활용하여 각 에이전트가 상기 복수의 도메인에 대해 관찰 및 액션을 수행하고 피드백 하는 멀티에이전트 강화학습을 통해 전체 특징 집합 중 동작 상상에 의미 있는 주요 특징을 추출하는 단계(S830), 주요 특징을 통해 상기 뇌파 데이터로부터 사용자의 움직임 의도를 판단함으로써 동작 상상을 분류하는 단계(S840)를 포함할 수 있다.
- [0125] 또한, 일 실시예에 따라서, 주요 특징을 추출하는 단계는, 공간-스펙트럼 에이전트를 통해 전체 특징 집합의 공간-스펙트럼 도메인 관찰을 수행하고, 정책을 수립하고, 액션을 결정 및 수행하는 공간-스펙트럼 특징 선택 단계, 시간 에이전트를 통해 전체 특징 집합의 시간 도메인 관찰을 수행하고, 정책을 수립하고, 액션을 결정 및 수행하는 시간 특징 선택 단계 및 공간-스펙트럼 에이전트 및 시간 에이전트의 액션 및 이에 따른 보상에 관한 정보를 받고, 공간-스펙트럼 에이전트 및 시간 에이전트와 정책 및 가치를 교환하며 피드백을 수행하는 중앙 집중형 크리틱 단계를 포함할 수 있다.
- [0126] 또한, 일 실시예에 따라서, 중앙 집중형 크리틱 단계는, $Q(s_t, a_t) = \mathbb{E}_{(s_{t+1:T}, a_{t+1:T})}[\mathcal{R}_t | s_t, a_t]$ (여기서 S_t 는 현재 스텝에서 주어진 특징 전체 집합에 대한 상태, a_t 는 공간-스펙트럼 에이전트와 시간 에이전트가 수행한 액션을 합친 공동 액션, R_t 는 현재 상태에서 공동 액션을 수행하였을 때 받게 될 보상의 총합, $\mathbb{E}$ 는 확률 변수에 대한 평균 값을 의미한다.)의 공동 액션 가치 함수에 기반하여 피드백을 수행하는 것을 특징으로 할 수 있다.
- [0127] 또한, 일 실시예에 따라서, 중앙 집중형 크리틱 단계는,

$$A^i(s_t, \mathbf{a}_t) = Q(s_t, \mathbf{a}_t; \psi) - \sum_{\tilde{\mathbf{a}}_t^i} \pi_t^i(\tilde{\mathbf{a}}_t^i | o_t^i; \theta^i) Q(s_t, (\mathbf{a}_t^{-i}, \tilde{\mathbf{a}}_t^i); \psi)$$
 (여기서 ψ 는 업데이트가 진행되는 중앙집중형 크리티크의 파라미터, π_t^i 는 해당 에이전트 i 가 가지고 있는 정책, $\tilde{\mathbf{a}}_t^i$ 는 해당 에이전트 i 가 선택할 수 있는 다른 종류의 액션, o_t^i 는 에이전트 i 의 관찰, θ^i 는 에이전트 i 의 파라미터, $\mathbf{a}_t^{-i}$ 는 해당 에이전트 i 를 제외한 다른 에이전트의 모든 액션을 고정시키는 것을 의미한다.)의 가치 함수를 상기 공간-스펙트럼 에이전트 및 상기 시간 에이전트에 할당되도록 하는 것을 특징으로 할 수 있다.

[0128] 또한, 일 실시예에 따라서, 중앙 집중형 크리티크 단계는, 가치 함수를 최대화하는 방향으로 파라미터 θ^i 를 업데이트하는 것을 특징으로 할 수 있다.

[0129] 이상 멀티에이전트 강화학습을 활용한 동작 상상 뇌파 분류 장치 및 방법의 여러 실시예에 대해 설명하였으나, 동작 상상 뇌파 분류 장치 및 방법은 오직 상술한 실시예에 한정되는 것은 아니다. 해당 기술 분야에서 통상의 지식을 가진 자가 상술한 실시예를 기초로 수정 및 변형하여 구현할 수 있는 다른 다양한 장치나 방법 역시 상술한 동작 상상 뇌파 분류 장치 및 방법의 일 실시예가 될 수 있다. 예를 들어, 설명된 방법(들)이 설명된 바와 다른 순서로 수행되거나, 및/또는 설명된 시스템, 구조, 장치, 회로 등의 구성 요소(들)가 설명된 바와 다른 형태로 결합, 연결 또는 조합되거나 다른 구성 요소 또는 균등물 등에 의하여 대체 또는 치환되더라도, 상술한 동작 상상 뇌파 분류 장치 및/또는 방법의 일 실시예가 될 수 있다.

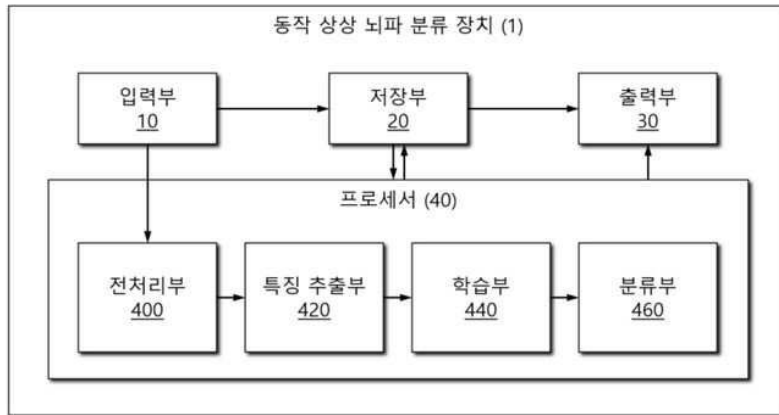
[0130] 본원 발명의 실시예 등과 관련된 기술 분야에서 통상의 지식을 가진 자는 상기 기재의 본질적인 특성에서 벗어나지 않는 범위에서 변형된 형태로 구현될 수 있음을 이해할 수 있을 것이다. 그러므로, 개시된 방법들은 한정적인 관점이 아닌 설명적 관점에서 고려되어야 한다. 본 발명의 범위는 발명의 상세한 설명이 아닌 특허청구 범위에 나타나며, 그와 동등한 범위 내에 있는 모든 차이점은 본 발명의 범위에 포함되는 것으로 해석되어야 한다.

부호의 설명

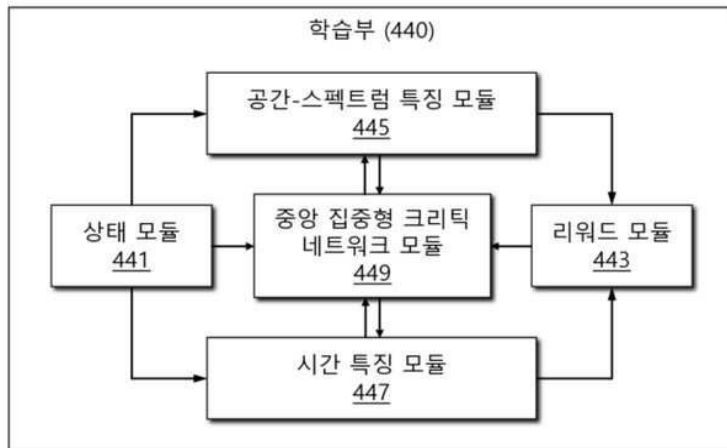
[0131] 1: 동작 상상 뇌파 분류 장치
10: 입력부
20: 저장부
30: 출력부
40: 프로세서
400: 전처리부
420: 특징 추출부
440: 학습부
441: 상태 모듈
443: 리워드 모듈
445: 공간-스펙트럼 특징 모듈
447: 시간 특징 모듈
449: 중앙 집중형 크리티크 네트워크 모듈
460: 분류부

도면

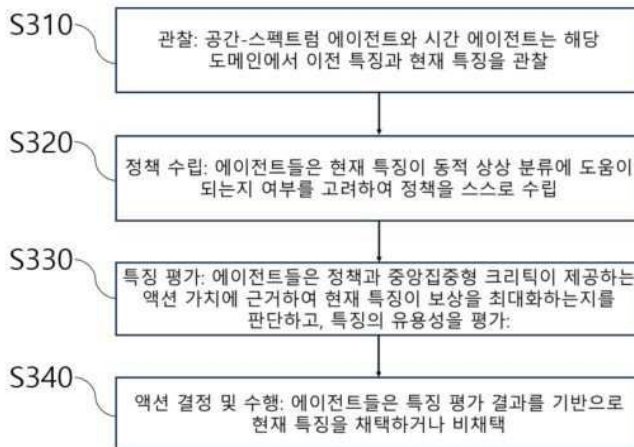
도면1



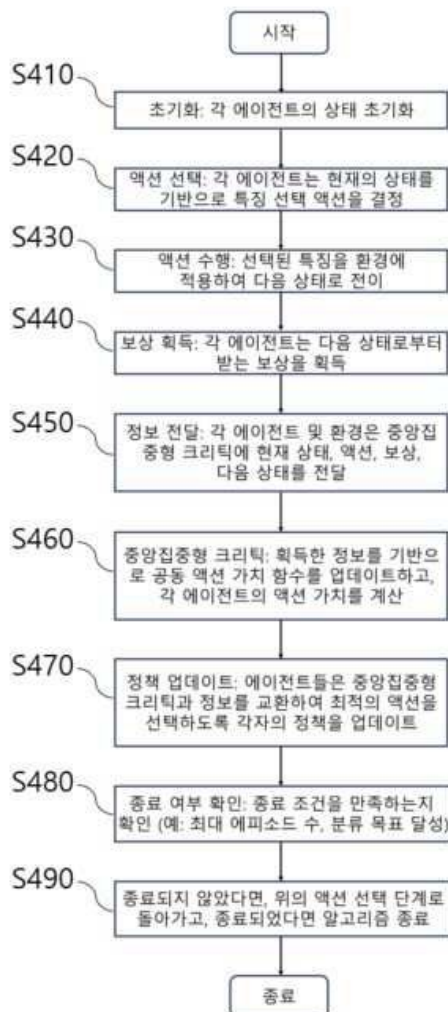
도면2



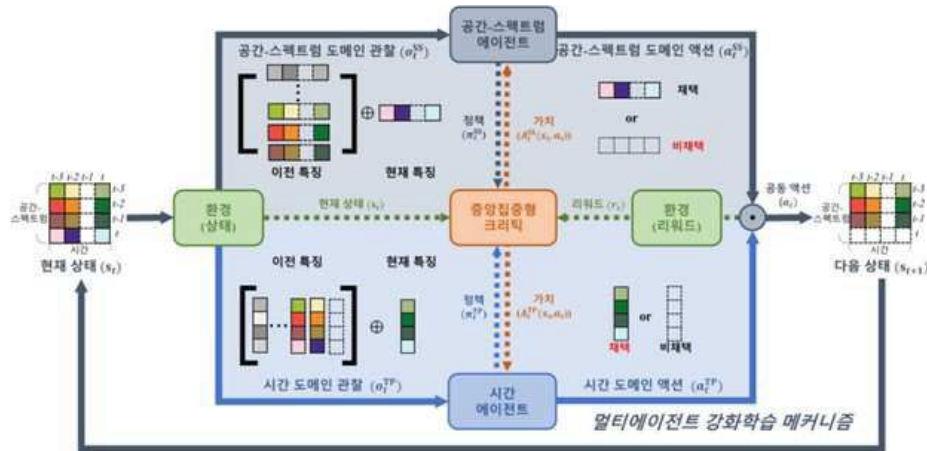
도면3



도면4



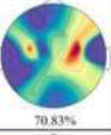
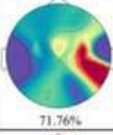
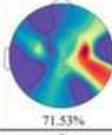
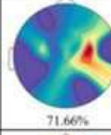
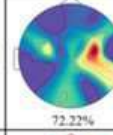
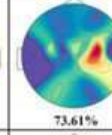
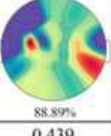
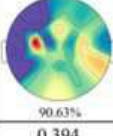
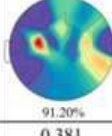
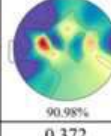
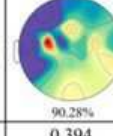
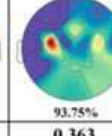
도면5



도면6

모델	특징 선택 도메인	상호작용 유무	분류 정확도 (데이터셋 1)	분류 정확도 (데이터셋 2)
EEGNet (특징 추출기)	X	X	69.59 ± 0.61	64.99 ± 0.20
+ SAT (단일 에이전트)	시간	X	70.51 ± 0.37	66.00 ± 0.23
+ SAS (단일 에이전트)	공간-스펙트럼	X	70.80 ± 0.95	66.04 ± 0.44
+ SAT & SAS (복수의 단일 에이전트)	시간, 공간-스펙트럼	X	70.69 ± 0.60	66.02 ± 0.30
+ SAS & SAT (복수의 단일 에이전트)	시간, 공간-스펙트럼	X	70.79 ± 0.40	66.04 ± 0.43
+ 본 발명 (멀티에이전트)	시간, 공간-스펙트럼	O	72.13 ± 0.62	66.74 ± 0.31

도면7

Method type	Baseline	Single-agent RL		Multiple single-agent RL		Multi-agent RL
Model	EEGNet	+ SAT	+ SAS	+ SAT&SAS	+ SAS&SAT	+ MARS (Ours)
(a) Left hand	 70.83%	 71.76%	 71.53%	 71.66%	 72.22%	 73.61%
(b) Right hand	 88.89%	 90.63%	 91.20%	 90.98%	 90.28%	 93.75%
Similarity	0.439	0.394	0.381	0.372	0.394	0.363

도면8

