

(19) 대한민국특허청(KR)
(12) 공개특허공보(A)(11) 공개번호 10-2023-0094409
(43) 공개일자 2023년06월28일

(51) 국제특허분류(Int. Cl.)

A61B 5/16 (2006.01) A61B 5/024 (2006.01)
A61B 5/0533 (2021.01) A61B 5/372 (2021.01)
A61B 5/397 (2021.01) A61B 5/398 (2021.01)
G16H 50/20 (2018.01)

(52) CPC특허분류

A61B 5/165 (2013.01)
A61B 5/02416 (2013.01)

(21) 출원번호 10-2021-0183580

(22) 출원일자 2021년12월21일

심사청구일자 2021년12월21일

(71) 출원인

인하대학교 산학협력단

인천광역시 미추홀구 인하로 100(용현동, 인하대 학교)

(72) 발명자

김덕환

서울특별시 양천구 목동로 230, 103동 1004호(목동, 아이파크아파트)

황우현

인천광역시 미추홀구 소성로 120, 115동 501호(학익동, 동아, 풍림아파트)

강동현

경상북도 포항시 북구 중흥로139번길 23-1, 1층(죽도동)

(74) 대리인

양성보

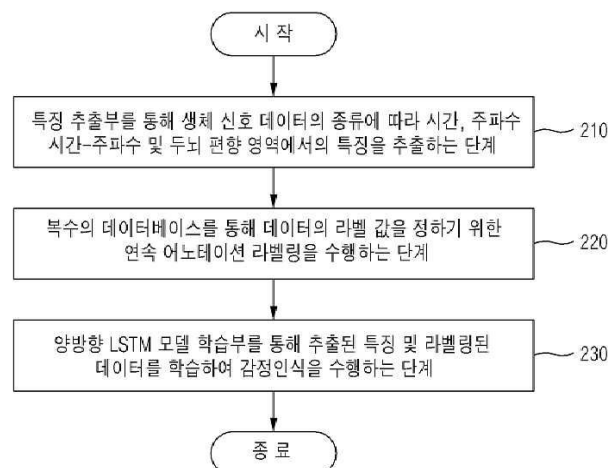
전체 청구항 수 : 총 8 항

(54) 발명의 명칭 감정인식을 위한 두뇌편향화 특징 추출 및 개미집단 최적화 기반 양방향 LSTM 신경망 모델 방법 및 장치

(57) 요약

감정인식을 위한 두뇌편향화 특징 추출 및 개미집단 최적화 기반 양방향 LSTM 신경망 모델을 이용한 감정인식 방법 및 장치가 제시된다. 본 발명에서 제안하는 감정인식을 위한 두뇌편향화 특징 추출 및 개미집단 최적화 기반 양방향 LSTM 신경망 모델을 이용한 감정인식 방법은 특징 추출부를 통해 생체 신호 데이터의 종류에 따라 시간, 주파수, 시간-주파수 및 두뇌 편향 영역에서의 특징을 추출하는 단계, 복수의 데이터베이스를 통해 데이터의 라벨 값을 정하기 위한 연속 어노테이션 라벨링을 수행하는 단계 및 양방향 LSTM 모델 학습부를 통해 추출된 특징 및 라벨링된 데이터를 학습하여 감정인식을 수행하는 단계를 포함한다.

대 표 도 - 도2



(52) CPC특허분류

A61B 5/0533 (2021.01)
A61B 5/372 (2022.01)
A61B 5/397 (2021.01)
A61B 5/398 (2022.01)
A61B 5/726 (2013.01)
A61B 5/7264 (2013.01)
G06N 3/049 (2023.01)
G06N 3/084 (2023.01)
G16H 50/20 (2018.01)

이 발명을 지원한 국가연구개발사업

과제고유번호	1711136767
과제번호	2021R1F1A1050750
부처명	과학기술정보통신부
과제관리(전문)기관명	한국연구재단
연구사업명	기본연구
연구과제명	[Ezbaro] IoT 데이터를 위한 연합 심층학습을 통한 블록체인지향 분산 엣지 컴퓨팅
솔루션에 대한 연구	
기 여 율	40/100
과제수행기관명	인하대학교
연구기간	2021.06.01 ~ 2022.02.28

이 발명을 지원한 국가연구개발사업

과제고유번호	1415171641
과제번호	N0002428
부처명	산업통상자원부
과제관리(전문)기관명	한국산업기술진흥원
연구사업명	미래형자동차 RND 전문인력양성사업
연구과제명	미래형자동차 R&D 전문인력양성 사업 (5차년도)
기 여 율	30/100
과제수행기관명	인하대학교
연구기간	2021.03.01 ~ 2022.02.28

이 발명을 지원한 국가연구개발사업

과제고유번호	1345332336
과제번호	2018R1D1A1B07042602
부처명	교육부
과제관리(전문)기관명	한국연구재단
연구사업명	기본연구(교육부)
연구과제명	[Ezbaro] 고가용성, 고신뢰성 클라우드 시스템을 위해 딥러닝 기술을 적용한
Key-Value 기반의 지능형 소프트웨어 정의 스토리지	
기 여 율	30/100
과제수행기관명	인하대학교
연구기간	2021.03.01 ~ 2021.05.31

명세서

청구범위

청구항 1

특징 추출부를 통해 생체 신호 데이터의 종류에 따라 시간, 주파수, 시간-주파수 및 두뇌 편향 영역에서의 특징을 추출하는 단계;

복수의 데이터베이스를 통해 데이터의 라벨 값을 정하기 위한 연속 어노테이션 라벨링(Continuous Annotation Labeling)을 수행하는 단계; 및

양방향 LSTM 모델 학습부를 통해 상기 추출된 특징 및 라벨링된 데이터를 학습하여 감정인식을 수행하는 단계를 포함하는 감정인식 방법.

청구항 2

제1항에 있어서,

상기 특징 추출부를 통해 생체 신호 데이터의 종류에 따라 시간, 주파수, 시간-주파수 및 두뇌 편향 영역에서의 특징을 추출하는 단계는,

시간 영역의 특징은 감정인식을 위해 시간에 따라 변화하는 생체 신호로부터 유도된 특징을 추출하고,

주파수 영역의 특징은 서로 다른 주파수 대역에서 나타나는 서로 다른 감정 상태를 출력하기 위해 PSD(Power spectrum density)를 슬로우 알파, 알파, 베타, 감마의 영역으로 추출하고,

시간-주파수 영역의 특징은 시간을 기준으로 복수의 비트로 분해하는 이산 웨이블릿 변환을 이용하고, 상기 이산 웨이블릿 변환으로 구해진 값에서 재귀 에너지 효율을 계산하며,

두뇌 편향 영역의 특징은 좌측 특정 감정이나 인지과정에서 활성화되는 뇌의 특정부위의 활성도를 이용하여 감정상태에 따른 좌뇌와 우뇌의 상관관계 특징을 이용하는

감정인식 방법.

청구항 3

제1항에 있어서,

상기 복수의 데이터베이스를 통해 데이터의 라벨 값을 정하기 위한 연속 어노테이션 라벨링을 수행하는 단계는,

MAHNOB-HCI 데이터베이스 및 MERTI-Apps 데이터베이스를 이용하고,

상기 MAHNOB-HCI 데이터베이스의 경우, FEELTRACE 프로그램 및 조이스틱을 사용하여 밸런스에 대하여 라벨링을 수행하고,

상기 MERTI-Apps 데이터베이스의 경우 MATLAB 프로그램을 통해 라벨링을 수행하는

감정인식 방법.

청구항 4

제1항에 있어서,

상기 양방향 LSTM 모델 학습부를 통해 상기 추출된 특징 및 라벨링된 데이터를 학습하여 감정인식을 수행하는 단계는,

셀상태와 게이트를 사용하는 LSTM 모델 학습을 수행하고,

상기 셀상태는 히든 레이어를 사용하며, 상기 히든 레이어를 갱신하기 위해 포갯(Forget), 입력(Input) 및 출력(Output) 게이트를 사용하며,

상기 포갯 게이트는 이전 상태의 히든 레이어의 값을 기억하는 정도를 결정하고, 상기 입력 게이트는 이전 상태에서 들어온 값이 아닌 새로운 값을 현재 히든 레이어에 반영할 정도를 결정하며, 상기 출력 게이트는 현재 셀 상태에서 출력할 값을 결정하고,

상기 양방향 LSTM 모델 학습을 위해 각각 정방향과 역방향으로 동작하는 두 개의 LSTM 모델 학습을 수행하는 감정인식 방법.

청구항 5

생체 신호 데이터의 종류에 따라 시간, 주파수, 시간-주파수 및 두뇌 편향 영역에서의 특징을 추출하는 특징 추출부;

데이터의 라벨 값을 정하기 위한 연속 어노테이션 라벨링(Continuous Annotation Labeling)을 수행하는 복수의 데이터베이스; 및

상기 추출된 특징 및 라벨링된 데이터를 학습하여 감정인식을 수행하는 양방향 LSTM 모델 학습부를 포함하는 감정인식 장치.

청구항 6

제5항에 있어서,

상기 특징 추출부는,

시간 영역의 특징은 감정인식을 위해 시간에 따라 변화하는 생체 신호로부터 유도된 특징을 추출하고,

주파수 영역의 특징은 서로 다른 주파수 대역에서 나타나는 서로 다른 감정 상태를 출력하기 위해 PSD(Power spectrum density)를 슬로우 알파, 알파, 베타, 감마의 영역으로 추출하고,

시간-주파수 영역의 특징은 시간을 기준으로 복수의 비트로 분해하는 이산 웨이블릿 변환을 이용하고, 상기 이산 웨이블릿 변환으로 구해진 값에서 재귀 에너지 효율을 계산하며,

두뇌 편향 영역의 특징은 화는 특정 감정이나 인지과정에서 활성화되는 뇌의 특정부위의 활성도를 이용하여 감정상태에 따른 좌뇌와 우뇌의 상관관계 특징을 이용하는

감정인식 장치.

청구항 7

제5항에 있어서,

상기 복수의 데이터베이스는,

MAHNOB-HCI 데이터베이스 및 MERTI-Apps 데이터베이스를 포함하고,

상기 MAHNOB-HCI 데이터베이스의 경우, FEELTRACE 프로그램 및 조이스틱을 사용하여 밸런스에 대하여 라벨링을 수행하고,

MERTI-Apps 데이터베이스의 경우 MATLAB 프로그램을 통해 라벨링을 수행하는

감정인식 장치.

청구항 8

제5항에 있어서,

상기 양방향 LSTM 모델 학습부는,

셀상태와 게이트를 사용하는 LSTM 모델 학습을 수행하고,

상기 셀상태는 히든 레이어를 사용하며, 상기 히든 레이어를 갱신하기 위해 포갯(Forget), 입력(Input) 및 출력(Output) 게이트를 사용하며,

상기 포갯 게이트는 이전 상태의 히든 레이어의 값을 기억하는 정도를 결정하고, 상기 입력 게이트는 이전 상태

에서 들어온 값이 아닌 새로운 값을 현재 히든 레이어에 반영할 정도를 결정하며, 상기 출력 게이트는 현재 셀 상태에서 출력할 값을 결정하고,

상기 양방향 LSTM 모델 학습을 위해 각각 정방향과 역방향으로 동작하는 두 개의 LSTM 모델 학습을 수행하는 감정인식 장치.

발명의 설명

기술 분야

[0001] 본 발명은 감정인식을 위한 두뇌편향화 특징 추출 및 개미집단 최적화 기반 양방향 LSTM 신경망 모델 기법 및 장치에 관한 것이다.

배경 기술

[0002] 최근 IoT 분야와 같은 다양한 로봇 인터페이스 기술의 발전으로 스마트 시계, AI 스피커와 같은 로봇의 역할이 인간의 생활에서 활용도가 높아지고 있다. 이에 따라 인공지능에 대한 중요성 또한 커지면서 인간과 로봇 간의 상호작용에 대한 연구도 많아졌다. 인간과 로봇 간의 상호작용을 위해서는 로봇이 인간의 의도 및 상태를 파악할 수 있어야 하고 딥러닝의 발전에 따라 인간의 내면 인식을 파악하는 기술에 대한 다양한 분야의 연구 또한 진행되었다. 인간의 감정 상태를 파악하는 연구에서는 이미지 데이터를 사용해 인간의 표정을 인식하거나 음성 데이터를 기반으로 감정을 파악하는 방법이 있다. 앞서 언급한 방법은 인간이 체성 신경계를 사용하여 본인의 표정과 목소리를 조절할 수 있기 때문에 감정을 드러내고 싶지 않다면 숨길 수 있어 내면의 감정을 찾기 힘들다는 단점이 있다. 하지만 생체신호가 기반이 되는 감정 인식의 경우, 자율 신경계 신호가 주된 데이터기 때문에 인간이 의도적으로 조절할 수 없어 인간 본연의 감정을 파악할 수 있다. 따라서 현재 감정인식에 대한 연구는 중추 신경계 또는 말초 신경계를 기반으로 한 연구가 많이 진행되고 있다.

[0003] 생체신호를 기반으로 하는 인공지능 감정인식 연구는 의도적으로 감정을 숨길 수 없다는 장점이 있지만 생체신호의 특성상 개개인의 특성이 강하고 외부의 환경요인을 최대한 없앤 실험환경이 필요하며, 피험자의 컨디션에 따라 데이터의 질이 다르기 때문에 성능적인 면에서는 이미지나 음성 데이터에 비해서 좋지 않다. 이러한 생체신호 취득의 어려움 때문에 생체신호를 사용하는 감정인식 연구에서는 공개 데이터베이스를 많이 사용하고 있다. 현재 생체신호 기반 감정인식 연구에서 많이 사용하고 있는 데이터베이스로는 DEAP과 MAHNOB-HCI, Smali, SEED가 있다. 하지만 감정인식을 위해 수집하는 생체신호의 종류와 전극의 위치가 정형화 되어있지 않고, 피험자들의 문화적 환경차이가 있어 각각의 데이터베이스에는 차이가 있다.

발명의 내용

해결하려는 과제

[0004] 본 발명이 이루고자 하는 기술적 과제는 수집한 생체신호에서 감정인식에 유효한 특징을 추출하면서 적절한 라벨 값을 매김으로써 모델의 감정인식 성능을 향상시키기 위한 감정인식을 위한 두뇌편향화 특징 추출 및 개미집단 최적화 기반 양방향 LSTM 신경망 모델 기법 및 장치를 제공하는데 있다. 제안된 방법을 평가하기 위해 각각의 생체신호에서 감정인식에 유효한 특징을 추출하고, 양방향 LSTM 모델 학습을 사용하여 성능을 확인하며, 두 가지의 데이터베이스를 비교하기 위해 같은 포맷으로 변환하고자 한다.

과제의 해결 수단

[0005] 일 측면에 있어서, 본 발명에서 제안하는 감정인식을 위한 두뇌편향화 특징 추출 및 개미집단 최적화 기반 양방향 LSTM 신경망 모델을 이용한 감정인식 방법은 특징 추출부를 통해 생체 신호 데이터의 종류에 따라 시간, 주파수, 시간-주파수 및 두뇌 편향 영역에서의 특징을 추출하는 단계, 복수의 데이터베이스를 통해 데이터의 라벨 값을 정하기 위한 연속 어노테이션 라벨링(Continuous Annotation Labeling)을 수행하는 단계 및 양방향 LSTM 모델 학습부를 통해 상기 추출된 특징 및 라벨링된 데이터를 학습하여 감정인식을 수행하는 단계를 포함한다.

[0006] 상기 특징 추출부를 통해 생체 신호 데이터의 종류에 따라 시간, 주파수, 시간-주파수 및 두뇌 편향 영역에서의 특징을 추출하는 단계는 시간 영역의 특징은 감정인식을 위해 시간에 따라 변화하는 생체 신호로부터 유도된 특징을 추출하고, 주파수 영역의 특징은 서로 다른 주파수 대역에서 나타나는 서로 다른 감정 상태를 출력하기 위해 PSD(Power spectrum density)를 슬로우 알파, 알파, 베타, 감마의 영역으로 추출하고, 시간-주파수 영역의

특징은 시간을 기준으로 복수의 비트로 분해하는 이산 웨이브렛 변환을 이용하고, 상기 이산 웨이브렛 변환으로 구해진 값에서 채워 에너지 효율을 계산하며, 두뇌 편향 영역의 특징은 화는 특정 감정이나 인지과정에서 활성화되는 뇌의 특정부위의 활성도를 이용하여 감정상태에 따른 좌뇌와 우뇌의 상관관계 특징을 이용한다.

[0007] 상기 복수의 데이터베이스를 통해 데이터의 라벨 값을 정하기 위한 연속 어노테이션 라벨링을 수행하는 단계는 MAHNOB-HCI 데이터베이스 및 MERTI-Apps 데이터베이스를 이용하고, 상기 MAHNOB-HCI 데이터베이스의 경우, FEELTRACE 프로그램 및 조이스틱을 사용하여 밸런스에 대하여 라벨링을 수행하고, 상기 MERTI-Apps 데이터베이스의 경우 MATLAB 프로그램을 통해 라벨링을 수행한다.

[0008] 상기 양방향 LSTM 모델 학습부를 통해 상기 추출된 특징 및 라벨링된 데이터를 학습하여 감정인식을 수행하는 단계는 셀상태와 게이트를 사용하는 LSTM 모델 학습을 수행하고, 상기 셀상태는 히든 레이어를 사용하며, 상기 히든 레이어를 갱신하기 위해 포갯(Forget), 입력(Input) 및 출력(Output) 게이트를 사용하며, 상기 포갯 게이트는 이전 상태의 히든 레이어의 값을 기억하는 정도를 결정하고, 상기 입력 게이트는 이전 상태에서 들어온 값이 아닌 새로운 값을 현재 히든 레이어에 반영할 정도를 결정하며, 상기 출력 게이트는 현재 셀상태에서 출력할 값을 결정하고, 상기 양방향 LSTM 모델 학습을 위해 각각 정방향과 역방향으로 동작하는 두 개의 LSTM 모델 학습을 수행한다.

[0009] 또 다른 일 측면에 있어서, 본 발명에서 제안하는 감정인식을 위한 두뇌편향화 특징 추출 및 개미집단 최적화 기반 양방향 LSTM 신경망 모델을 이용한 감정인식 장치는 생체 신호 데이터의 종류에 따라 시간, 주파수, 시간-주파수 및 두뇌 편향 영역에서의 특징을 추출하는 특징 추출부, 데이터의 라벨 값을 정하기 위한 연속 어노테이션 라벨링(Continuous Annotation Labeling)을 수행하는 복수의 데이터베이스 및 상기 추출된 특징 및 라벨링된 데이터를 학습하여 감정인식을 수행하는 양방향 LSTM 모델 학습부를 포함한다.

발명의 효과

[0010] 본 발명의 실시예들에 따른 감정인식을 위한 두뇌편향화 특징 추출 및 개미집단 최적화 기반 양방향 LSTM 신경망 모델 기법 및 장치를 통해 수집한 생체신호에서 감정인식에 유효한 특징을 추출하면서 적절한 라벨 값을 매김으로써 모델의 감정인식 성능을 향상시킬 수 있다. 제안된 방법을 평가하기 위해 각각의 생체신호에서 감정인식에 유효한 특징을 추출하고, 양방향 LSTM 모델 학습을 사용하여 성능을 확인하며, 두 가지의 데이터베이스를 비교하기 위해 같은 포맷으로 변환한다.

도면의 간단한 설명

[0011] 도 1은 본 발명의 일 실시예에 따른 본 발명의 일 실시예에 따른 생체 신호 유형을 나타내는 도면이다.

도 2는 본 발명의 일 실시예에 따른 감정인식을 위한 두뇌편향화 특징 추출 및 개미집단 최적화 기반 양방향 LSTM 신경망 모델을 이용한 감정인식 방법을 설명하기 위한 흐름도이다.

도 3은 본 발명의 일 실시예에 따른 감정인식을 위한 두뇌편향화 특징 추출 및 개미집단 최적화 기반 양방향 LSTM 신경망 모델을 이용한 감정인식 장치의 구성을 나타내는 도면이다.

도 4는 본 발명의 일 실시예에 따른 MAHNOB-HCI 데이터베이스 및 MERTI-Apps 데이터베이스를 설명하기 위한 도면이다.

도 5는 본 발명의 일 실시예에 따른 LSTM 모델 학습과정을 설명하기 위한 도면이다.

도 6은 본 발명의 일 실시예에 따른 양방향 LSTM 모델 학습부를 설명하기 위한 도면이다.

발명을 실시하기 위한 구체적인 내용

[0012] 인간의 내면 인식에 대한 기술은 인간과 로봇사이의 상호작용을 위한 핵심적 기술이다. 그 중 감정인식은 의도 인지 연구와 더불어 인간의 내면 상태를 인식하기 위한 대표적인 연구로 자리매김 되었다.

[0013] 본 발명에서는 말초 신경계 신호 PPG, GSR, EMG, EOG와 중추 신경계 신호 EEG를 사용한다. 본 발명에서 제안하는 모델은 말초 신경계 신호를 사용하여 말초 신경계 신호의 유효한 특징과 중추 신경계에서 두뇌편향화를 통한 특징을 추출해 양방향 LSTM 네트워크 모델의 입력 데이터로 적용한 새로운 생체신호 감정 인식 방식이다. 신호의 특징은 시간, 주파수, 시간-주파수 영역에서 추출하였다. 감정의 분류는 어노테이션 라벨링(Annotation Labeling)을 수행하여 밸런스(Valence)와 어로우절(Arousal)을 기준으로 포지티브(Positive), 네거티브

(Negative)로 나누어진다. 그리고 라벨링 데이터(Labeling data)의 값은 -1과 1 사이의 값이다.

[0014]

본 발명에 따른 모델에 사용되는 생체신호 데이터의 경우 자체 실험을 통해 수집한 데이터베이스를 사용하여 모델을 학습하지만 모델의 신뢰성을 높이고 데이터베이스의 성능을 비교하기 위해 MAHNOB-HCI 데이터베이스도 사용하여 모델의 성능을 확인하였다. 앞서 언급한 생체신호의 단점으로 인해 음성, 영상과 비교하여 생체신호를 사용한 경우의 성능은 비교적 낮다. 따라서 최근 생체신호의 연구에서는 이러한 단점을 보완하기 위해 SVM, k-NN과 같은 기계학습부터 CNN, LSTM같은 심층학습까지 다양한 학습 모델을 적용하였다. 실험 결과 자체제작 데이터셋을 사용했을 때는 RMSE 0.0598, MAHNOB-HCI 데이터셋을 사용했을 때 RMSE 0.0482를 보였다. 이하, 본 발명의 실시 예를 첨부된 도면을 참조하여 상세하게 설명한다.

[0016]

감정에 대한 연구는 심리학분야에서 처음으로 시작되었다. 하지만 현재에서는 심리학뿐만 아니라 공학분야에서도 다양한 연구가 진행되고 있다. 그 중 감정인식에 대한 연구는 인간과 로봇의 상호작용을 목표로 다양한 분야에서 진행되고 있다. 하지만 오랫동안 지속되어 온 연구에도 불구하고 감정의 분류에 대한 정확한 기준이 없고 감정의 표현에 대한 방법이 문화와 환경에 따라 다르기 때문에 높은 분류 정확도를 달성하기 어렵다. 이러한 감정인식 연구에서 사용하는 대표적인 데이터로는 이미지, 음성, 텍스트, 생체신호 등이 있다. 먼저 이미지 데이터는 인간의 얼굴 표정을 통해 나타나는 데이터이다. 얼굴 표정은 인간이 감정을 표현하는데 있어 가장 알기 쉽고, 자연스러운 비언어적 의사소통 방법이다. 사람의 얼굴과 눈, 눈썹, 입을 기준으로 변화를 찾아내 전처리 단계를 거친 다음 특징으로 추출하고, 이를 기계학습 또는 심층학습 모델을 통해 감정인식을 수행한다[1]. 음성 데이터는 인간의 의사소통 방법 중에서 가장 빠르고 자연스러운 방법이기 때문에 감정인식 분야에서 가장 많이 사용된 연구 분야이다. 그리고 이러한 특징은 인간과 인간 사이뿐만 아니라 인간과 로봇 간의 상호작용에서도 효과적으로 발현된다. 앞서 언급한 3가지 대표적인 데이터 중에서 음성인식 분야가 실생활에서 가장 많이 상용화되어 있다. 하지만 이러한 발전에도 불구하고 음성을 통해 로봇이 인간의 감정 상태를 인식하는 것은 어려운 이유가 있다. 첫 번째로 감정을 구별하는데 어떤 음성 기능이 가장 유용한지 명확하지 않다는 점이다. 두 번째로는 말하는 사람의 문화와 환경에 따라 감정이 표현되는 방법이 다르기 때문에 이러한 기준을 세우는 것이 어렵다는 점이다. 그래서 대부분의 음성데이터를 사용하는 감정인식 연구에서는 실험을 진행함에 있어서 피험자들 간의 문화적 차이가 없는 것으로 가정을 한다[2]. 텍스트의 경우, 특정 감정의 키워드 또는 어휘 친화성을 통해 감정인식을 분류한다. 텍스트를 이용한 감정인식 연구는 표준 감정 키워드의 계층이 없기 때문에 인지 심리학에서 연구된 감정 시스템을 기반으로 감정을 분류한다[3]. 앞서 언급한 3가지 연구 분야들은 전부 사람의 겉모습이나 의도적인 표현을 통해 외면 감정을 인식하기 때문에 만약 피험자가 실제 느낀 내면의 감정을 숨기고 다른 감정을 표현한다면 충분히 속일 수 있다. 또한 같은 감정을 느끼더라도 사람마다 무표정, 높낮이가 없는 목소리 등의 표현적 차이가 있기 때문에 감정을 분류하는데 있어 한계점이 있다. 생체신호를 사용한 감정인식 연구분야에서 로(raw) 데이터를 전처리하는 대표적인 방법이 두 가지가 있다. 먼저 로 데이터를 사용하여 생체신호 데이터를 이미지화 시켜 모델을 학습시키는 방법과 로 데이터에서 각각의 생체신호에서의 감정인식에 유용한 적절한 파형의 특징을 추출하여 모델을 학습시키는 방법이다. 두 가지의 방법은 모두 장점이 있으나 현재 더 많이 연구하는 방법은 특징을 추출하는 방법이다.

[0017]

감정을 분류하는 기준은 다양하지만 현재 연구에서 가장 많이 사용하는 방법은 제임스 러셀(James Russell)의 감정 원형 모델이다. 이 모델은 각성과 원자가 차원을 사용하여 2차원 원형 공간을 만든다. 각성은 수직 축, 원자가는 수평 축을 나타내며, 원의 중심은 중립 수준의 각성과 원자가를 나타낸다. 이 모델에서 감정 상태는 모든 수준의 원자가 및 각성과 하나 또는 둘 다의 중립 수준을 표현할 수 있다.

[0018]

도 1은 본 발명의 일 실시예에 따른 생체 신호 유형을 나타내는 도면이다.

[0019]

감정인식 분야에서 많은 연구가 진행되고 학습 모델이 발전됨에 따라 생체신호의 연구 또한 발전되었다. 현재 감정인식 분야에서 사용되고 있는 신호는 EEG 뿐만 아니라 PPG, ECG, EMG, EOG, GSR 등이 있다. 이 신호들 중에서 EEG 신호만 중추 신경계 신호이며, 나머지 신호들은 말초 신경계 신호이다. 중추 신경계는 뇌와 척수로 구분되며 신경세포와 신경섬유가 집합하고 상호 연계하여 중심부를 형성하는 부분이다. 말초 신경계는 외부 자극을 중추 신경계로 전달하고 전신의 각 기관에게 중추 신경계에서 파생된 신호를 전달해주는 신경계이며, 2가지로 구분된다. 말초 신경계에는 체성 신경계와 자율 신경계가 있는데 체성 신경계는 EMG, EOG가 포함되며, 자율 신경계에는 PPG, ECG, GSR가 포함된다. 종래기술에서는 EEG 단일 신호를 사용해서 감정인식을 하였으나, EEG 단일 신호를 사용했을 때의 성능이 매우 낮기 때문에 최근에는 EEG 뿐만 아니라 말초 신경계 계통의 생체신호를 같이 취득하여 연구에 사용하고 있다. ECG, PPG신호는 심장에서 파생되는 신호로 피험자의 긴장 또는 분노의 감정을

잘 찾을 수 있으며, GSR 신호는 손의 땀과 체온 변화를 측정하고 그에 따른 피부 긴장도를 파악한다. 따라서 피험자가 스트레스를 받거나 예상치 못한 상황에 따른 긴장 상태를 파악할 수 있다. 하지만 GSR신호는 사람마다 땀을 흘리는 양이 다르기 때문에 절대적인 기준이 없어 사람마다 기준이 달라지게 된다. 따라서 실험 환경에서 최대한 신체 활동을 줄이고 쾌적한 환경을 조성해야 한다. EMG와 EOG 신호는 각각 얼굴의 근육 신호와 안구 신호이다. 얼굴에서 파생된 신호이기 때문에 감정인식뿐만 아니라 다양한 인간과 로봇사이의 상관관계 연구에서 많이 사용되고 있는 신호이다. EMG 신호는 근육세포막에서 발생하는 전기 신호이며, EOG신호는 각막을 양극, 망막을 음극의 형태로 미약한 전기가 발생되어 생기는 신호이다. EOG는 일반적으로 수평 EOG와 수직 EOG로 나누어진다. 이러한 점을 이용하여 안구의 움직임에 대한 각도와 속도 같은 자세한 정보를 얻을 수 있어 피험자의 시선이 어디에 있는지 찾는 안구 추적 기술을 만들 수 있다. EMG와 EOG신호는 얼굴에서 파생되는 신호이기 때문에 이미지 기반 감정인식 기술과 비슷한 지표를 가진다. 마지막으로 EEG신호는 뇌에서 발생하는 전위차에 의해 발생하는 뇌전류를 두피 상으로 유도하여 취득한다. 따라서 다른 생체신호 수집장비와는 다른 특별한 수집장비가 필요하다. 특히EEG신호가 다른 생체신호와 다른 점은 주파수마다 독특한 특징을 가진다는 점이다. 그리고 두피의 위치마다 특정 주파수의 높고 낮음이 다르다. 예를 들어, EEG 신호의 가장 대표적인 파형인 알파파는 10KHz 전후의 규칙적인 신호로서 눈을 감고 평온한 상태에서 나타난다. 또한 두정엽과 후두엽에서 가장 크고 전두엽에서 가장 작게 나타난다. 만약 피험자가 눈을 뜨고 물체를 주시하거나 정신적으로 흥분하게 되면 알파파가 억제되고 불규칙한 파동이 발생한다. 알파파보다 빠른 베타파형은 중앙엽과 전두엽에서 크게 나타난다. 이외에도 감마, 세타, 슬로우 알파 파형이 있으며, 이러한 독특한 특징을 가진 뇌파를 사용해 감정인식뿐만 아니라 간질, 뇌종양, 뇌 손상 등의 질환도 진단할 수 있다.

- [0021] 도 2는 본 발명의 일 실시예에 따른 감정인식을 위한 두뇌편향화 특징 추출 및 개미집단 최적화 기반 양방향 LSTM 신경망 모델을 이용한 감정인식 방법을 설명하기 위한 흐름도이다.
- [0022] 본 발명에서는 생체신호를 통한 감정인식에서 가장 많이 활용되고 있는 중추신경계 EEG신호를 일반적인 특징뿐만 아니라 두뇌편향화를 통한 특징 그리고 말초신경계에서 추출한 특징을 사용하여 양방향 LSTM 모델에 적용하여 감정인식의 정확도를 높이는 방법을 제안한다. 중추신경계 신호인 뇌신호는 발현되는 밸런스(Valence) 감정에 따라 활성화되는 두뇌의 위치가 다르고, 말초신경계 신호에서는 피험자의 어로우절(Arousal) 감정인 긴장상태와 스트레스 정도를 인식하는 것에 높은 인식률을 보이기 때문에 양방향 LSTM 모델을 사용하는 것으로 감정분류의 정확도를 향상시킬 수 있다. 감정분류는 자체 제작한 데이터베이스를 통해 밸런스와 어로우절을 기반으로 연속 어노테이션 라벨링(Continuous Annotation Labeling)을 진행하였다. 그리고 MAHNOB-HCI 데이터베이스와 비교하여 제안하는 모델과 자체 제작 데이터베이스의 성능이 효과적임을 확인할 수 있었다.
- [0023] 제안하는 감정인식을 위한 두뇌편향화 특징 추출 및 개미집단 최적화 기반 양방향 LSTM 신경망 모델을 이용한 감정인식 방법은 특징 추출부를 통해 생체 신호 데이터의 종류에 따라 시간, 주파수, 시간-주파수 및 두뇌 편향 영역에서의 특징을 추출하는 단계(210), 복수의 데이터베이스를 통해 데이터의 라벨 값을 정하기 위한 연속 어노테이션 라벨링(Continuous Annotation Labeling)을 수행하는 단계(220) 및 양방향 LSTM 모델 학습부를 통해 상기 추출된 특징 및 라벨링된 데이터를 학습하여 감정인식을 수행하는 단계(230)를 포함한다.
- [0024] 단계(210)에서, 특징 추출부를 통해 생체 신호 데이터의 종류에 따라 시간, 주파수, 시간-주파수 및 두뇌 편향 영역에서의 특징을 추출한다.
- [0025] 본 발명의 실시예에 따른 시간 영역의 특징은 감정인식을 위해 시간에 따라 변화하는 생체 신호로부터 유도된 특징을 추출할 수 있다.
- [0026] 본 발명의 실시예에 따른 주파수 영역의 특징은 서로 다른 주파수 대역에서 나타나는 서로 다른 감정 상태를 출력하기 위해 PSD(Power spectrum density)를 슬로우 알파, 알파, 베타, 감마의 영역으로 추출할 수 있다.
- [0027] 본 발명의 실시예에 따른 시간-주파수 영역의 특징은 시간을 기준으로 복수의 비트로 분해하는 이산 웨이브렛 변환을 이용하고, 상기 이산 웨이브렛 변환으로 구해진 값에서 재귀 에너지 효율을 계산할 수 있다.
- [0028] 본 발명의 실시예에 따른 두뇌 편향 영역의 특징은 화는 특정 감정이나 인지과정에서 활성화되는 뇌의 특정부위의 활성도를 이용하여 감정상태에 따른 좌뇌와 우뇌의 상관관계 특징을 이용할 수 있다.
- [0029] 단계(220)에서, 복수의 데이터베이스를 통해 데이터의 라벨 값을 정하기 위한 연속 어노테이션 라벨링(Continuous Annotation Labeling)을 수행한다.

- [0030] 본 발명의 실시예에 따르면 복수의 데이터베이스는 MAHNOB-HCI 데이터베이스 및 MERTI-Apps 데이터베이스를 포함할 수 있다.
- [0031] 본 발명의 실시예에 따른 MAHNOB-HCI 데이터베이스의 경우, FEELTRACE 프로그램 및 조이스틱을 사용하여 밸런스에 대하여 라벨링을 수행한다.
- [0032] 본 발명의 실시예에 따른 MERTI-Apps 데이터베이스의 경우 MATLAB 프로그램을 통해 라벨링을 수행한다.
- [0033] 단계(230)에서, 양방향 LSTM 모델 학습부를 통해 상기 추출된 특징 및 라벨링된 데이터를 학습하여 감정인식을 수행한다.
- [0034] 본 발명의 실시예에 따른 양방향 LSTM 모델 학습부는 셀상태와 게이트를 사용하는 LSTM 모델 학습을 수행할 수 있다. 상기 셀상태는 히든 레이어를 사용하며, 상기 히든 레이어를 갱신하기 위해 포겟(Forget), 입력(Input) 및 출력(Output) 게이트를 사용할 수 있다.
- [0035] 본 발명의 실시예에 따른 포겟 게이트는 이전 상태의 히든 레이어의 값을 기억하는 정도를 결정할 수 있다.
- [0036] 본 발명의 실시예에 따른 입력 게이트는 이전 상태에서 들어온 값이 아닌 새로운 값을 현재 히든 레이어에 반영할 정도를 결정할 수 있다.
- [0037] 본 발명의 실시예에 따른 출력 게이트는 현재 셀상태에서 출력할 값을 결정할 수 있다.
- [0038] 본 발명의 실시예에 따른 양방향 LSTM 모델 학습부는 양방향 LSTM 모델 학습을 위해 각각 정방향과 역방향으로 동작하는 두 개의 LSTM 모델 학습을 수행한다.
- [0040] 도 3은 본 발명의 일 실시예에 따른 감정인식을 위한 두뇌편향화 특징 추출 및 개미집단 최적화 기반 양방향 LSTM 신경망 모델을 이용한 감정인식 장치의 구성을 나타내는 도면이다.
- [0041] 본 발명에서는 수집한 생체신호에서 감정인식에 유효한 특징을 추출하면서 적절한 라벨 값을 매김으로써 모델의 감정인식 성능을 향상시키는 것을 목표로 한다. 제안된 양방향 LSTM 신경망을 평가하기 위해 각각의 생체신호에서 감정인식에 유효한 특징을 추출하고, 양방향 LSTM모델을 사용하여 성능을 확인한다. 두 가지의 데이터베이스를 비교하기 위해 같은 포맷으로 변환하고, MANHOB-HCI 데이터베이스의 ECG는 MERTI-Apps의 PPG로 대체한다. 본 발명의 실시예에 따른 실험에서 입력은 10-fold를 사용하고 RMSE로 결과를 확인하였다.
- [0042] 제안하는 감정인식을 위한 두뇌편향화 특징 추출 및 개미집단 최적화 기반 양방향 LSTM 신경망 모델을 이용한 감정인식 장치(300)는 특징 추출부(310), 복수의 데이터베이스(320) 및 양방향 LSTM 모델 학습부(330)를 포함한다.
- [0043] 본 발명의 실시예에 따른 특징 추출부(310)는 생체 신호 데이터의 종류에 따라 시간, 주파수, 시간-주파수 및 두뇌 편향 영역에서의 특징을 추출한다.
- [0044] 본 발명의 실시예에 따른 시간 영역의 특징은 감정인식을 위해 시간에 따라 변화하는 생체 신호로부터 유도된 특징을 추출할 수 있다.
- [0045] 본 발명의 실시예에 따른 주파수 영역의 특징은 서로 다른 주파수 대역에서 나타나는 서로 다른 감정 상태를 출력하기 위해 PSD(Power spectrum density)를 슬로우 알파, 알파, 베타, 감마의 영역으로 추출할 수 있다.
- [0046] 본 발명의 실시예에 따른 시간-주파수 영역의 특징은 시간을 기준으로 복수의 비트로 분해하는 이산 웨이블릿 변환을 이용하고, 상기 이산 웨이블릿 변환으로 구해진 값에서 제귀 에너지 효율을 계산할 수 있다.
- [0047] 본 발명의 실시예에 따른 두뇌 편향 영역의 특징은 화는 특정 감정이나 인지과정에서 활성화되는 뇌의 특정부위의 활성도를 이용하여 감정상태에 따른 좌뇌와 우뇌의 상관관계 특징을 이용할 수 있다. 이하, 본 발명의 실시예에 따른 특징 추출부(310)에 관하여 더욱 상세하게 설명한다.
- [0048] 본 발명의 실시예에 따른 모델 학습을 위해 선택한 특징은 생체 신호 데이터의 종류에 따라 다르다. 먼저 추출된 특징은 시간, 주파수, 시간-주파수 영역에서 MATLAB을 통해 추출하였다. 각 데이터베이스에서 추출한 특징의 개수는 37개로 동일하지만 EEG 채널의 수가 다르므로 전체 특징 데이터의 크기도 다르다. MAHNOB-HCI 데이터베이스의 경우 32x37로 1184개이며, MERTI-Apps데이터베이스의 경우 12x37로 444개이다. 다른 생체신호들과 다르게 EEG는 슬로우 알파, 알파, 베타, 감마 총 4가지 대역으로 구분하여 특징을 추출하였다.

[0049] 먼저, 본 발명의 실시예에 따른 시간 영역의 특징은 감정인식을 위해 시간에 따라 변화하는 EEG 신호로부터 유도된 특징을 추출한다. 기존에는 EEG의 공간분해능 유리함 때문에 주파수 영역에서만 특징을 사용하였지만, 본 발명에서는 최종 특징 집단을 구성하기 때문에 포함되었다. 평균(mean), 최소(minimum), 최대(maximum), 힘(power), 표준편차(standard deviation), 1차 차이(1st difference), 정규화 된 1차 차이(normalized 1st difference), 2차 차이(2st difference), 정규화 된 2차 차이(normalized 2st difference)의 특징은 서로 다른 감정을 분류하는 기능을 목적으로 한다. Hjorth 특징은 본 발명에서 제안하는 활동 영역, 이동성 및 복잡도를 다루는 매개변수이다. 이 매개변수는 EEG 신호의 평균 출력과 평균 빈도 및 기울기 수를 나타낸다. 프랙탈 차원(fractal dimension)의 특성을 계산하는 방법에 대한 여러 기법 중에서 Higuchi 방법은 종래기술의 Box-counting 기법보다 좋은 성능을 보였다. 마지막 특징은 EEG 신호의 비선형성의 정도를 포착하기 위하여 로켓 평균 변화를 추정하는 NSI(Non-Stationary Index)이다.

[0050] 본 발명의 실시예에 따른 주파수 영역 특징은 기존 연구에서 가장 많이 사용된 특징이다. 서로 다른 주파수 대역에서 나타나는 출력은 서로 다른 감정 상태를 알아내기 좋은 요소이다. 본 발명의 실시예에 따르면, PSD(Power spectrum density)를 슬로우 알파(Slow Alpha, 8-10Hz), 알파(Alpha, 8-12.9Hz), 베타(Beta, 13-29.9Hz), 감마(Gamma, 30-50Hz) 4가지 영역으로 추출한다.

[0051] 본 발명의 실시예에 따른 주파수 영역 특징은 시간 영역에 비해 정확도가 높게 분류가 되는 특징이지만 시간적 설명이 부족하다. 때문에 본 발명의 실시예에 따른 시간-주파수 영역 특징은 주파수 영역과 같이 4가지의 주파수 영역으로 5가지의 특징을 추출한다. 가장 자주 쓰이는 특징은 이산 웨이블릿 변환(Discrete Wavelet Transform; DWT)이다. DWT는 신호를 시간을 기준으로 하여 다양한 비트로 분해한다. 음성분야에서도 사용되며 감정인식 분야에도 사용된다. 이외에 DWT로 구해진 값에서 재귀 에너지 효율(Recursive Energy Efficiency; REE)이 계산된다.

[0052] 본 발명의 실시예에 따른 두뇌 편향화는 특정 감정이나 인지 과정에서 뇌의 특정부위 활성화가 전문화 될 수 있다는 이론이다. 보통 두뇌 편향화는 좌뇌와 우뇌로 구분하여 나눈다. 알려진 사실로는 시야의 왼쪽 부분이 우뇌에서 처리되고, 오른쪽 부분이 좌뇌에서 처리된다는 것이다. 이처럼 시야뿐만 아니라 감정상태도 좌뇌와 우뇌에 따라 구분될 수 있다. 그렇기 때문에 본 발명에서는 감정상태에 따른 좌뇌와 우뇌의 상관관계 특징을 이용하였다.

$$B_t = \xi_{rl} \circ \frac{(\zeta_l - \zeta_r)}{(\zeta_l + \zeta_r)} \quad (1)$$

[0054] 여기서, B_t 는 뇌의 비대칭 특징이다. 이 값을 얻기 위해서 행렬 ξ_{rl} 과 ζ_l , ζ_r 사이에서 하다마드 곱을 수행한다. 행렬 ξ_{rl} 은 왼쪽과 오른쪽 전극 채널 간의 인과성 비대칭 행렬이며, ζ_l , ζ_r 행렬은 각각 왼쪽 반구와 오른쪽 반구의 특정 대역의 스펙트럼 파워의 대수이다. 행렬 ξ_{rl} 를 계산하는 식은 다음과 같다:

$$\xi_{rl}(m, n) = \frac{|A_{rl}(m, n)|}{\sqrt{a_k^H(m, n) a_k(m, n)}} \quad (2)$$

[0056] 이 수식은 그랜저(Granger) 인과 관계의 개념을 기반으로 하는 PDC(partial Directed Coherence) 측정 수식이다. m 과 n 의 값은 각각 1부터 m 과 1부터 n 까지다. A_{rl} 은 $A(m, n)$ 의 rl 번째 요소이며, a_k^H 은 a_k 의 켈레전치 행렬이다. A_{rl} 을 위한 $A(m, n)$ 의 수식은 다음과 같다:

$$A(m, n) = I - \sum_{d=1}^p A_d(n) z^{-d} \Big|_{z=e^{j2\pi f}} \quad (3)$$

[0058] I 는 단위 행렬이며, 주파수 m 의 범위는 0부터 나이퀴스트 율까지이다.

$$A_d = \begin{bmatrix} a_{11}^d & \cdots & a_{1M}^d \\ \vdots & \ddots & \vdots \\ a_{M1}^d & \cdots & a_{MM}^d \end{bmatrix} \quad (4)$$

이는 MVAR 모델을 사용하여 계산하는데 GDC(Granger Causality Index) 및 DTF(Directed Transfer Function) 및 PDC와 같은 그랜저 인과 관계 기반 측정법이다. 4번 행렬의 각각의 요소는 지연 d에서의 채널 r과 l사이의 선형 관계를 의미한다. 이를 통해서 왼쪽과 오른쪽 EEG 채널 간의 영향을 확인할 수 있다.

다시 도 3을 참조하면, 본 발명의 실시예에 따른 복수의 데이터베이스(320)는 데이터의 라벨 값을 정하기 위한 연속 어노테이션 라벨링(Continuous Annotation Labeling)을 수행한다.

본 발명의 실시예에 따른 복수의 데이터베이스(320)는 MAHNOB-HCI 데이터베이스(321) 및 MERTI-Apps 데이터베이스(322)를 포함할 수 있다.

본 발명의 실시예에 따른 MAHNOB-HCI 데이터베이스(321)의 경우, FEELTRACE 프로그램 및 조이스틱을 사용하여 밸런스에 대하여 라벨링을 수행한다.

본 발명의 실시예에 따른 MERTI-Apps 데이터베이스(322)의 경우 MATLAB 프로그램을 통해 라벨링을 수행한다.

도 4는 본 발명의 일 실시예에 따른 MAHNOB-HCI 데이터베이스 및 MERTI-Apps 데이터베이스를 설명하기 위한 도면이다.

도 4(a)는 MAHNOB-HCI 데이터베이스를 설명하기 위한 도면이다. 도 4(b)는 MERTI-Apps 데이터베이스를 설명하기 위한 도면이다.

도 4(a)를 참조하면, 본 발명의 실시예에 따른 MAHNOB-HCI 데이터베이스는 27명의 피험자를 대상으로 데이터를 수집하였으며, 데이터는 EEG와 ECG, GSR, EMG, 호흡 패턴, 피부 온도 등의 PNS, 신체 비디오, 시야 추적, 오디오를 수집하였다. MAHNOB-HCI 데이터베이스는 역겨움, 즐거움, 행복함, 무서움, 슬픔, 중립 6가지의 감정으로 분류되고 평균 길이 81.4초의 짧은 20개의 비디오를 선별하여 피험자에게 시청하게 함으로써 실험을 진행하였다. 그리고 영상이 종료될 때마다 주어진 평가지에 영상에 대한 감정 평가를 작성하게 하였다. 평가지는 감정 키워드, 각성, 원자가, 우세, 예측 가능성 등 5가지 주제에 대해 작성하였으며 감정 키워드를 제외한 주제의 척도는 1 ~ 9 사이의 값으로 기준을 정하였다. 모든 생체신호는 256Hz 로 수집되었으며, ECG와 GSR 신호에서는 시간적 저주파 드리프트를 제거하였다. GSR장비는 가운데 손가락과 집게 손가락에 전극을 배치하였으며, ECG장비는 3 개의 전극을 양쪽 쇄골 뼈 아래와 복부에 배치하였다. 피부 온도 장비는 새끼 손가락에 온도 센서를 배치하여 기록하였으며, 호흡 패턴 장비는 복부 주위에 호흡 벨트를 착용하여 측정되었다. EEG 신호 수집 장비는 국제 10-20 시스템에 따라 배치된 활성 AgCl 32개의 전극이 있는 캡 장비를 사용하였다. 사용된 EEG의 주파수는 세타, 슬로우 알파, 알파, 베타, 감마 파형이며, 32개의 전극과 6개의 파형을 통해 총 216개의 특징을 추출하였다. 그리고 말초 신경계 신호에서는 총 102개의 특징을 추출하였다. 각 데이터의 라벨 값은 원자가를 기준으로 어노테이션 라벨링이 진행되었다.

도 4(b)를 참조하면, MERTI-Apps 데이터베이스는 여자 34명, 남자 28명의 총 62명, 평균 나이 23.95의 피험자들을 대상으로 데이터를 수집하였으며, 수집한 데이터는 EEG, EOG, GSR, PPG 신호이다. 실험에 사용된 비디오는 총 5개이며, 유도한 감정은 혐오, 행복, 슬픔, 분노, 즐거움이다. 비디오의 길이는 60초-206초이며, 평균 114.6초의 짧은 비디오다. 감정은 지속될수록 익숙해지고 반응이 약해지므로 비교적 짧지만 강하게 감정이 유발될 수 있는 비디오들을 선정하였다. 그리고 유발되었던 감정이 다음 비디오에 영향을 주면 안되기 때문에 비디오와 비디오 사이에 흑백화면을 제공함으로써 피험자가 다음 비디오를 볼 수 있는 준비 기간을 제공한다.

본 발명의 실시예에 따른 생체신호 수집장비는 BIOPACK M150 장비를 사용하였으며, EEG의 경우 국제 10-20 시스템에 의거한 12개의 전극에서 수집하였고, GSR은 가운데 손가락과 약지손가락에 전극을 위치시키고 PPG는 검지 손가락에 전극을 위치시켜서 데이터를 수집하였다. EOG 신호는 수직 EOG 신호를 취득하였다. 왜냐하면 EEG신호는 매우 민감하기 때문에 눈 깜빡임에 의한 인공물에도 영향을 받는다. 따라서 이 인공물을 제거하기 위해 EOG 신호를 취득하여 사용한다. 결과적으로 수집된 데이터는 총 310개이고, 각 데이터의 라벨 값은 각성과 원자가가

각각 -100 ~ 100 사이의 값으로 포함되어 있다.

- [0072] 본 발명의 실시예에 따른 MAHNOB-HCI 데이터베이스와 MERTI-Apps 데이터베이스는 데이터의 라벨 값을 정하기 위해 연속 어노테이션 라벨링을 수행한다. MAHNOB-HCI 데이터베이스의 경우, FEELTRACE라는 프로그램과 조이스틱을 사용하여 밸런스에 대하여 라벨링을 수행하였다.
- [0073] 본 발명의 실시예에 따른 MERTI-Apps 데이터베이스의 경우 라벨링 프로그램은 자체적으로 MATLAB 프로그램을 통해 제작하였다. 그리고 객관적인 어노테이션 평가를 위해 실험과 전혀 무관한 관찰자들을 5명 선정하여 본격적인 어노테이션 라벨링을 진행하기 전에 몇 번의 훈련을 거쳐 객관적인 라벨링 값이 나올 수 있도록 하였다. 어노테이션 라벨링의 진행방식은 관찰자가 피험자의 정면 영상을 보고 피험자의 표정에 따라 각성과 원자의 점수를 측정한다. 이때 피험자의 정면 영상의 재생속도는 0.5배속이다. 그 이유는 관찰자가 피험자의 감정표현을 보다 더 잘 파악할 수 있기 때문이다. 라벨링의 점수는 -100 ~ 100의 범위를 가지고 있고, 프로그램에 피험자의 번호와 실험 영상의 번호를 입력하면 셀프-평가(Self-Assessment) 박스에 피험자가 자체평가에서 기록한 결과가 출력되어 관찰자들에게 피험자가 어떠한 감정을 느꼈는지에 대한 힌트를 제공한다. 본 발명의 실시예에 따른 실험에서는 생체신호뿐만 아니라 압력센서를 설치하여 피험자가 감정의 강도를 강하게 느끼고 있다고 생각할 경우 누르게 하였다. 이는 피험자가 해당 영상에 대해 얼마만큼의 감정을 느꼈고, 강하게 느꼈는지에 대한 척도를 시각적으로 볼 수 있게 해준다. 본 발명의 실시예에 따른 어노테이션 라벨링은 관찰자의 결과 데이터의 품질을 향상시키기 위해 피험자의 정면 영상을 0.5배속으로 하여 진행하였다.
- [0075] 다시 도 3을 참조하면, 본 발명의 실시예에 따른 양방향 LSTM 모델 학습부(330)는 상기 추출된 특징 및 라벨링된 데이터를 학습하여 감정인식을 수행한다.
- [0076] 본 발명의 실시예에 따른 양방향 LSTM 모델 학습부(330)는 셀상태와 게이트를 사용하는 LSTM 모델 학습을 수행할 수 있다. 상기 셀상태는 히든 레이어를 사용하며, 상기 히든 레이어를 갱신하기 위해 포겟(Forget), 입력(Input) 및 출력(Output) 게이트를 사용할 수 있다.
- [0077] 본 발명의 실시예에 따른 포겟 게이트는 이전 상태의 히든 레이어의 값을 기억하는 정도를 결정할 수 있다.
- [0078] 본 발명의 실시예에 따른 입력 게이트는 이전 상태에서 들어온 값이 아닌 새로운 값을 현재 히든 레이어에 반영할 정도를 결정할 수 있다.
- [0079] 본 발명의 실시예에 따른 출력 게이트는 현재 셀상태에서 출력할 값을 결정할 수 있다.
- [0080] 본 발명의 실시예에 따른 양방향 LSTM 모델 학습부는 양방향 LSTM 모델 학습을 위해 각각 정방향과 역방향으로 동작하는 두 개의 LSTM 모델 학습을 수행한다.
- [0082] 도 5는 본 발명의 일 실시예에 따른 LSTM 모델 학습과정을 설명하기 위한 도면이다.
- [0083] LSTM은 RNN의 한 부분으로서 지속적인 데이터를 분석하는데 사용된다. 기존의 RNN은 구조적으로 바나싱 그레이디언트(vanishing gradients)의 문제가 있어 단계가 길어지면 길어질수록 오래된 정보가 없어지게 되어, 현재의 값과 오래 전에 발생한 값 사이를 분석할 수 없게 된다. 이 문제를 해결하기 위한 모델이 도 5의 LSTM 모델이다. 이러한 RNN의 단점을 LSTM모델은 셀상태와 게이트를 사용하여 극복하였다. 이러한 해결 방법을 사용하여 LSTM모델은 이전 데이터의 값을 메모리의 형태로 기억할 수 있게 되었고 시간이 길어져도 이전 정보가 없어지지 않게 되었다. 셀상태는 기본적으로 히든 레이어(Hidden Layer)로 생각되며, 히든 레이어를 갱신하기 위해 LSTM에서는 포겟(Forget), 입력(Input), 출력(Output) 3가지의 게이트를 사용한다. 먼저 포겟 게이트는 이전 시간의 히든 레이어들의 값을 어느 정도 기억하는지 결정하는 역할을 한다. 값은 0과 1로 나누어지며 0일 경우는 이전 셀의 모든 값을 잊고 1일 경우는 모두 기억하게 된다. 그리고 입력 게이트는 이전 상태에서 들어온 값이 아닌 새로운 값을 현재 히든 레이어에 얼마만큼 반영할지를 결정한다. 출력 게이트는 현재 셀상태에서 출력할 값을 결정한다. 일반적인 LSTM 모델은 이렇게 이전 시간들의 단계들이 다음 단계에 영향을 줄 것이라는 가정하에 있다. 하지만 양방향 LSTM의 경우는 이전 시간들의 상태뿐만 아니라 이후 시간들의 상태도 현재의 상태에 영향을 주게 된다.
- [0085] 도 6은 본 발명의 일 실시예에 따른 양방향 LSTM 모델 학습부를 설명하기 위한 도면이다.

[0086] 도 6과 같이 양방향 LSTM은 하나가 아닌 두 개의 LSTM을 이어서 사용한다. 두 개의 LSTM은 같은 방향이 아닌 각각 정방향(Forward)과 역방향(Backward)으로 이루어져 있다. 시간 간격이 $i-1$ 부터 $i+2$ 까지 있다고 가정하면 정방향 LSTM에서는 입력 데이터를 $i-1$ 부터 $i+2$ 까지 순차적으로 식(5)을 사용하여 학습을 한다. 반면에 역방향 LSTM은 입력 데이터를 $i+2$ 부터 $i-1$ 까지 식(6)을 사용하여 반대방향으로 학습을 한다. 이렇게 학습을 하면 각각의 시간마다 각자 다른 방향의 히든 레이어가 생성된다. 이러한 두 개의 히든 레이어는 학습된 가중치를 통해서 식(7)과 같이 하나의 히든 레이어인 y_t 로 합쳐진다.

$$\vec{h}_t = \sigma(W_{x\vec{h}}x_t + W_{\vec{h}\vec{h}}\vec{h}_{t-1} + b_{\vec{h}}) \quad (5)$$

$$\overleftarrow{h}_t = \sigma(W_{x\overleftarrow{h}}x_t + W_{\overleftarrow{h}\overleftarrow{h}}\overleftarrow{h}_{t-1} + b_{\overleftarrow{h}}) \quad (6)$$

$$y_t = W_{\vec{h}y}\vec{h}_t + W_{\overleftarrow{h}y}\overleftarrow{h}_t + b_y \quad (7)$$

[0090] 본 발명의 실시예에 따른 정방향 LSTM의 경우, 시간 t 를 1부터 T 까지 시간간격 마다 입력 데이터를 정방향으로 실행하여 모든 출력 데이터를 도출할 수 있다.

[0091] 각 단계마다 활성화 정보를 저장하고 히든 레이어에 정방향으로 이전 단계의 정보를 전달하고, 각각의 시간 간격 LSTM 셀에서 나온 출력 데이터를 정방향 히든 벡터 $\vec{h}_t$ 에 전달한다.

[0092] 본 발명의 실시예에 따른 역방향 LSTM의 경우, 시간 t 를 T 부터 1까지 시간간격 마다 입력 데이터를 역방향으로 실행하여 모든 출력 데이터를 도출할 수 있다.

[0093] 각 단계마다 활성화 정보를 저장하고 히든 레이어에 역방향으로 이전 단계의 정보를 전달하고, 각각의 시간 간격 LSTM 셀에서 나온 출력 데이터를 역방향 히든 벡터 $\overleftarrow{h}_t$ 에 전달한다.

[0094] 본 발명의 실시예에 따른 MAHNOB-HCI 데이터베이스와 자체 제작된 MERTI-Apps 데이터베이스를 사용하여 모델의 성능을 검증하였다. 각각의 데이터는 0.25초, 즉 4Hz마다 특징을 추출하였고, 모델을 사용할 때 타임스텝(Timestep)을 8로 설정해 2초의 데이터를 가지고 감정인식에 대한 학습을 순차적으로 진행하였다. 양방향 LSTM 모델의 성능 비교를 위해 기본적인 LSTM과 FC-LSTM 모델 또한 동일한 조건을 통해 비교를 수행하였다. 이후 MAHNOB-HCI 데이터베이스와 MERTI-Apps 데이터베이스 간의 성능 비교 또한 진행하였다. 학습모델은 Pycharm을 사용하여 구현하였으며, 실험 환경은 CPU는 Ryzen 7 2700X, GPU는 RTX 2080Ti를 사용하였다.

[0095] MAHNOB-HCI 데이터베이스는 앞서 언급했듯이 현재 생체신호를 사용하는 감정인식 연구에서 많이 사용되고 있는 공개 데이터베이스이다. MAHNOB-HCI 데이터베이스의 어노테이션 라벨 데이터는 밸런스 데이터로만 이루어져 있다. 먼저 기본적인 LSTM, FC-LSTM, 양방향 LSTM의 모델 성능 비교를 MAHNOB-HCI 데이터베이스를 사용하여 수행하였다. 전체 데이터를 10-fold cross validation으로 트레인 데이터 90%, 테스트 데이터는 10%로 구분했다. 그리고 트레인 데이터에서 세부적으로 학습 데이터 70%, 유효데이터(Validation) 데이터 20%로 구분했다. 성능 비교는 PSD 특징만 사용한 결과, 3 도메인 특징을 사용한 결과, 마지막으로 두뇌편향화 특징을 추가한 실험을 진행하였다.

[0096] <표 1>

Feature \ Model	Basic LSTM[5] [RMSE]	FC-LSTM [RMSE]	Bi-directional LSTM [RMSE]
PSD	0.053 ± 0.029	0.0523 ± 0.017	0.051 ± 0.01
3 Domain	0.0524 ± 0.039	0.0519 ± 0.022	0.0487 ± 0.024
3 Domain + Brain Lateralization	0.052 ± 0.028	0.0493 ± 0.023	0.0456 ± 0.019

[0097]

[0098] 표 1을 보면 실험결과 MANOB-HCI 데이터베이스에서 사용한 기본 LSTM의 성능이 가장 낮게 나왔으며, 본 발명에

서 제안하는 양방향 LSTM의 성능이 모든 특징부분에서 가장 높게 나왔다.

[0099] 본 발명의 실시예에 따른 MERTI-Apps 데이터베이스는 MAHNOB-HCI 데이터베이스의 실험을 참고하여 자체제작한 데이터베이스다. MAHNOB-HCI 데이터베이스와 차이점은 영상의 개수가 적고 피험자의 수가 많다는 점이다. 앞서 실험한 MAHNOB-HCI 데이터베이스의 양방향 LSTM 데이터와 비교를 하기 위해 3 도메인 특징과 두뇌 편향과 특징 두 가지에 대한 실험을 진행하였다. 실험 조건은 MAHNOB-HCI 데이터베이스 실험과 같은 조건으로 진행하였고, 전체 특징 데이터의 크기와 데이터의 라벨의 차이가 있었다. MAHNOB-HCI 데이터베이스의 경우 label 데이터는 밸런스에 대한 데이터만 존재하지만 MERTI-Apps 데이터베이스는 밸런스뿐만 아니라 어로우절 데이터 또한 존재한다. 따라서 MERTI-Apps의 특징 데이터 444개와 함께 라벨 데이터 2개가 추가된 Time x 446의 입력데이터가 된다. 실험결과 각 데이터베이스에 대한 RMSE값을 확인하였다. MAHNOB-HCI 데이터베이스 결과가 MERTI-Apps 데이터베이스보다 성능이 좋다는 것을 볼 수 있다. 이에 대한 원인으로 MAHNOB-HCI 데이터베이스는 피험자가 적고 영상이 많지만 MERTI-Apps 데이터베이스는 피험자가 많고 영상이 적다. 생체신호의 특성상 개인차가 있는 신호이므로 피험자가 다양할수록 컴퓨터의 학습능력이 낮아진다. 이와 더불어 MAHNOB-HCI 데이터베이스는 한사람당 20개의 영상에서 데이터를 추출하지만 MERTI-Apps 데이터베이스의 경우는 많은 사람에게서 5개의 영상을 추출하기 때문에 개인차가 더욱 심해지게 되기 때문이다.

[0100] 이와 같이, 본 발명에서는 다양한 생체신호를 기반으로 양방향 LSTM을 사용해 감정인식을 하는 방법에 대해 제안하였다. 결과적으로 과거와 미래의 데이터를 사용해 현재의 상태를 인식하는 방법은 생체신호 분야에서도 유효함을 확인할 수 있었다. 또한, 특징 추출단계에서 가장 많이 사용하는 3개의 영역과 별개로 두뇌편향화 영역이 감정인식에 효과적이라는 사실도 확인하였다. MAHNOB-HCI 데이터베이스에서 3개의 영역을 사용했을 때 RMSE 값이 0.0487의 값을 얻었지만 두뇌편향화 영역을 포함하여 실험을 한 결과 RMSE 0.0456을 얻을 수 있었다. MERTI-Apps 데이터베이스에서도 3개의 영역보다 4개의 영역을 사용한 결과 RMSE 값이 0.0053 감소한 것을 확인할 수 있었다. 두 데이터베이스의 실험결과를 비교하면 비교적 자체 제작한 MERTI-Apps 데이터베이스의 결과가 낮은 것을 확인할 수 있다. 이는 상술된 개인차와 데이터 양에 따른 결과이다.

[0101] 본 발명의 실시예에 따른 MERTI-Apps 데이터베이스는 5개의 영상을 사용하였다. 기존에 MERTI-Apps 데이터베이스 제작하기 위한 실험은 총 15개의 영상을 사용해 데이터를 수집하였다. 하지만 본 발명에서는 5개의 영상을 사용했다. 그 이유는 실험을 진행함에 있어 5개 영상마다 사용하는 생체신호 조합을 변화시키면서 수집하였기 때문에 CNS와 PNS를 같이 수집하는 단계가 5개 영상밖에 없었기 때문이다. 따라서 MERTI-Apps 데이터베이스의 성능 향상을 위해서는 실험을 진행하는데 있어 CNS와 PNS를 동시에 수집하는 단계에서 영상을 늘려 데이터 수집량을 늘려야 한다. 만약 추가적으로 이러한 사항을 개선한다면 MERTI-Apps 데이터베이스 또한 MAHNOB-HCI 데이터베이스만큼 성능 향상이 될 것으로 예상된다.

[0102] 본 발명의 실시예에 따른 양방향 LSTM 알고리즘을 통해 실험을 진행하면서 다양한 특징을 사용하여 감정인식 연구를 진행하였다. 감정인식 연구분야는 앞서 언급했듯이 인간과 컴퓨터 사이의 상호작용을 위해서 진행하고 있는 연구이다. 그렇기 때문에 학습모델 연구에서 추구하는 목표 중 하나는 실시간 모델이다. 실시간으로 인간의 감정을 인식하고 결과에 따라 컴퓨터가 반응을 표현하는 것이다. 하지만 본 발명의 실험에서는 Epoch를 50으로 설정하였고 각 데이터베이스 마다 적지 않은 양의 데이터가 있기 때문에 상당히 긴 학습시간이 소요되었다. 따라서 쓸모 없는 데이터는 사용하지 않고 효율적인 데이터만을 사용한다면 학습시간을 상당히 줄일 수 있다. 차원 축소 기술과 유효한 특징 선택 기술이 적용된다면 학습 시간 감소뿐만 아니라 더 좋은 성능의 결과를 도출할 수 있다. 또한, 생체신호만을 단일로 사용하지 않고 영상 또는 음성 데이터를 함께 사용하여 멀티모달로 데이터를 확장한다면 더 좋은 성능 향상을 얻을 수 있다.

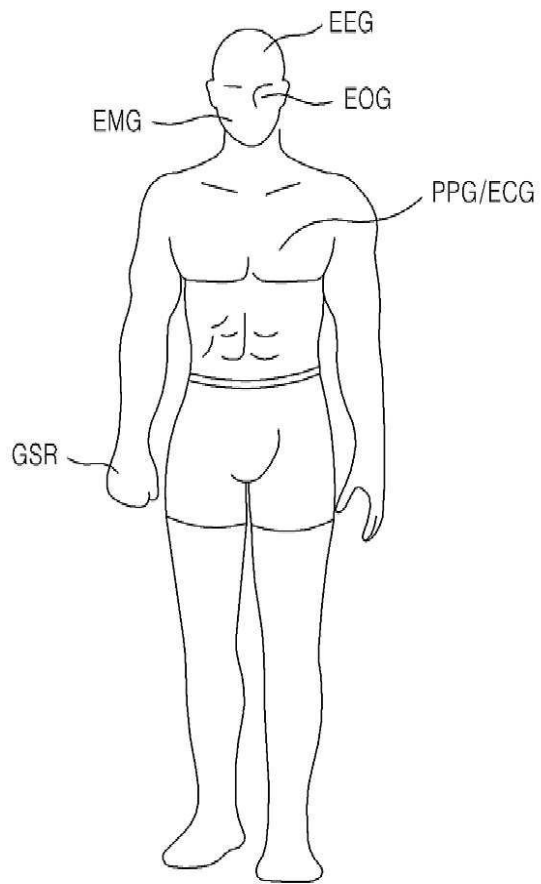
[0104] 이상에서 설명된 장치는 하드웨어 구성요소, 소프트웨어 구성요소, 및/또는 하드웨어 구성요소 및 소프트웨어 구성요소의 조합으로 구현될 수 있다. 예를 들어, 실시예들에서 설명된 장치 및 구성요소는, 예를 들어, 프로세서, 콘트롤러, ALU(arithmetic logic unit), 디지털 신호 프로세서(digital signal processor), 마이크로컴퓨터, FPA(field programmable array), PLU(programmable logic unit), 마이크로프로세서, 또는 명령(instruction)을 실행하고 응답할 수 있는 다른 어떠한 장치와 같이, 하나 이상의 범용 컴퓨터 또는 특수 목적 컴퓨터를 이용하여 구현될 수 있다. 처리 장치는 운영 체제(OS) 및 상기 운영 체제 상에서 수행되는 하나 이상의 소프트웨어 애플리케이션을 수행할 수 있다. 또한, 처리 장치는 소프트웨어의 실행에 응답하여, 데이터를 접근, 저장, 조작, 처리 및 생성할 수도 있다. 이해의 편의를 위하여, 처리 장치는 하나가 사용되는 것으로 설명된 경우도 있지만, 해당 기술분야에서 통상의 지식을 가진 자는, 처리 장치가 복수 개의 처리 요소

(processing element) 및/또는 복수 유형의 처리 요소를 포함할 수 있음을 알 수 있다. 예를 들어, 처리 장치는 복수 개의 프로세서 또는 하나의 프로세서 및 하나의 컨트롤러를 포함할 수 있다. 또한, 병렬 프로세서(parallel processor)와 같은, 다른 처리 구성(processing configuration)도 가능하다.

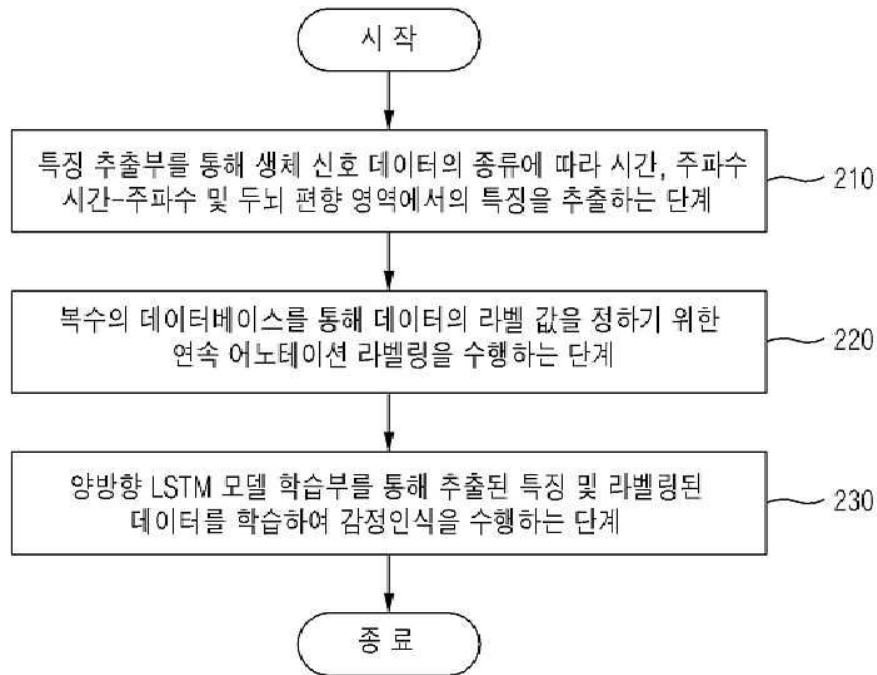
- [0105] 소프트웨어는 컴퓨터 프로그램(computer program), 코드(code), 명령(instruction), 또는 이들 중 하나 이상의 조합을 포함할 수 있으며, 원하는 대로 동작하도록 처리 장치를 구성하거나 독립적으로 또는 결합적으로(collectively) 처리 장치를 명령할 수 있다. 소프트웨어 및/또는 데이터는, 처리 장치에 의하여 해석되거나 처리 장치에 명령 또는 데이터를 제공하기 위하여, 어떤 유형의 기계, 구성요소(component), 물리적 장치, 가상 장치(virtual equipment), 컴퓨터 저장 매체 또는 장치에 구체화(embodiment)될 수 있다. 소프트웨어는 네트워크로 연결된 컴퓨터 시스템 상에 분산되어서, 분산된 방법으로 저장되거나 실행될 수도 있다. 소프트웨어 및 데이터는 하나 이상의 컴퓨터 판독 가능 기록 매체에 저장될 수 있다.
- [0106] 실시예에 따른 방법은 다양한 컴퓨터 수단을 통하여 수행될 수 있는 프로그램 명령 형태로 구현되어 컴퓨터 판독 가능 매체에 기록될 수 있다. 상기 컴퓨터 판독 가능 매체는 프로그램 명령, 데이터 파일, 데이터 구조 등을 단독으로 또는 조합하여 포함할 수 있다. 상기 매체에 기록되는 프로그램 명령은 실시예를 위하여 특별히 설계되고 구성된 것들이거나 컴퓨터 소프트웨어 당업자에게 공지되어 사용 가능한 것일 수도 있다. 컴퓨터 판독 가능 기록 매체의 예에는 하드 디스크, 플로피 디스크 및 자기 테이프와 같은 자기 매체(magnetic media), CD-ROM, DVD와 같은 광기록 매체(optical media), 플롭티컬 디스크(floptical disk)와 같은 자기-광 매체(magneto-optical media), 및 롬(ROM), 램(RAM), 플래시 메모리 등과 같은 프로그램 명령을 저장하고 수행하도록 특별히 구성된 하드웨어 장치가 포함된다. 프로그램 명령의 예에는 컴파일러에 의해 만들어지는 것과 같은 기계어 코드뿐만 아니라 인터프리터 등을 사용해서 컴퓨터에 의해서 실행될 수 있는 고급 언어 코드를 포함한다.
- [0107] 이상과 같이 실시예들이 비록 한정된 실시예와 도면에 의해 설명되었으나, 해당 기술분야에서 통상의 지식을 가진 자라면 상기의 기재로부터 다양한 수정 및 변형이 가능하다. 예를 들어, 설명된 기술들이 설명된 방법과 다른 순서로 수행되거나, 및/또는 설명된 시스템, 구조, 장치, 회로 등의 구성요소들이 설명된 방법과 다른 형태로 결합 또는 조합되거나, 다른 구성요소 또는 균등물에 의하여 대치되거나 치환되더라도 적절한 결과가 달성될 수 있다.
- [0108] 그러므로, 다른 구현들, 다른 실시예들 및 특허청구범위와 균등한 것들도 후술하는 특허청구범위의 범위에 속한다.
- [0109] <참고 문헌>
- [0110] [1] Widanagamaachchi, W. N., & Dharmaratne, A. T. (2009). Emotion Recognition with Image Processing and Neural Networks. Research Gate.
- [0111] [2] El Ayadi, M., Kamel, M. S., & Karray, F. (2011). Survey on speech emotion recognition: Features, classification schemes, and databases. Pattern Recognition, 44(3), 572-587.
- [0112] [3] Shivhare, S. N., & Khethawat, S. (2012). Emotion detection from text. arXiv preprint arXiv:1205.4944.
- [0113] [4] B. H. Kim, and S. Jo (2018). Deep Physiological Affect Network for the Recognition of Human Emotions. IEEE Transactions on Affective Computing. early access
- [0114] [5] Plutchik, R. (1980): Emotion: A psychoevolutionary synthesis. Harpercollins College Division
- [0115] [6] Soleymani, M., Asghari-Esfeden, S., Fu, Y., & Pantic, M. (2015). Analysis of EEG signals and facial expressions for continuous emotion detection. IEEE Transactions on Affective Computing, 7(1), 17-28.

도면

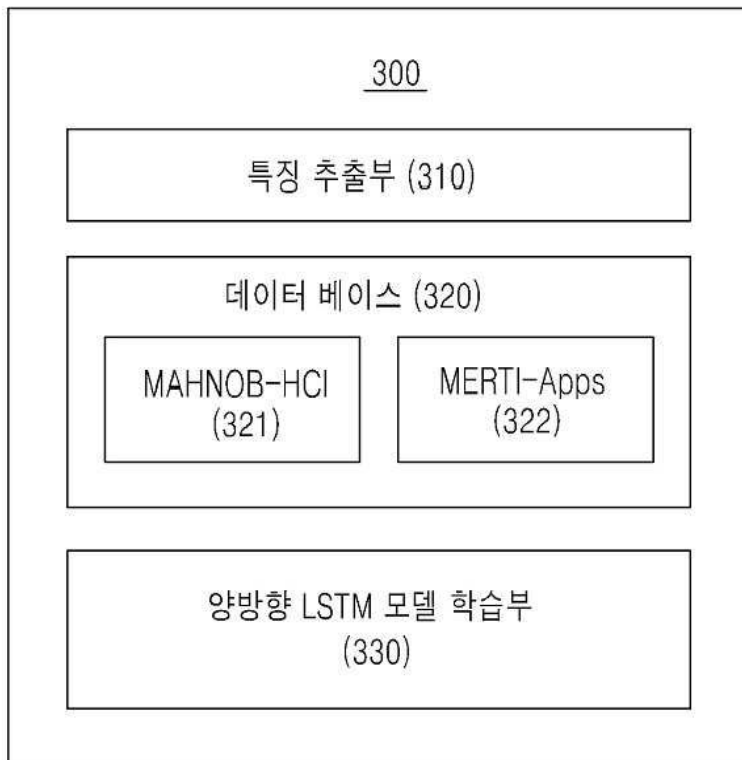
도면1



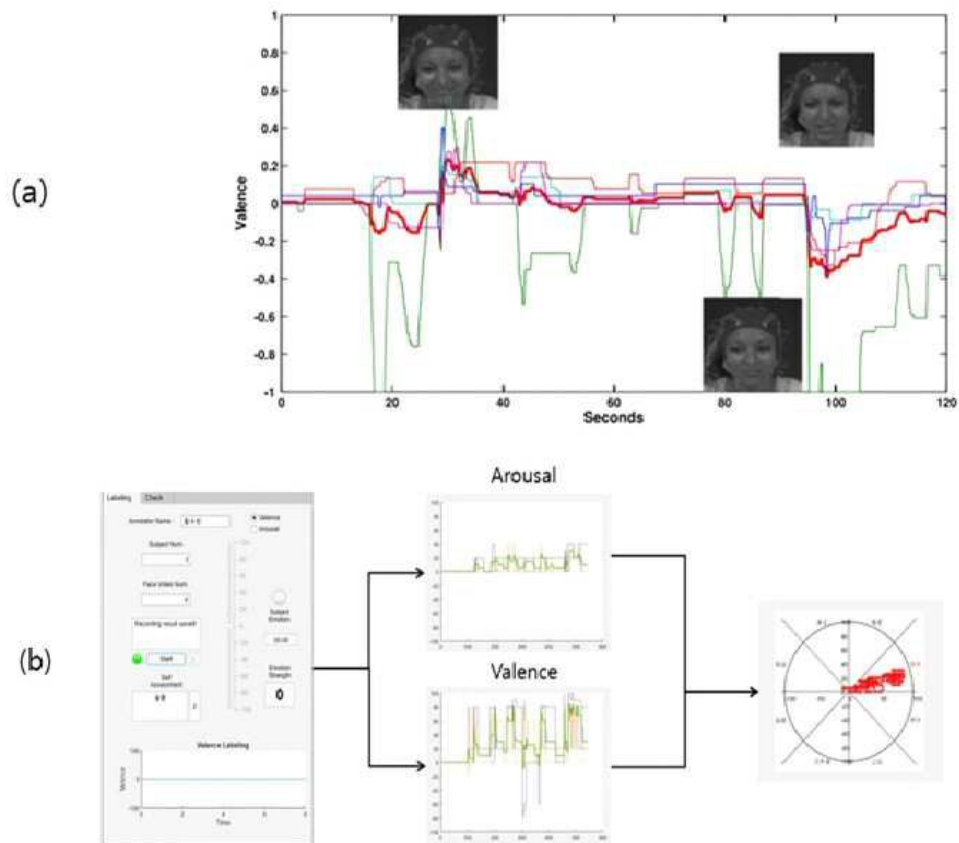
도면2



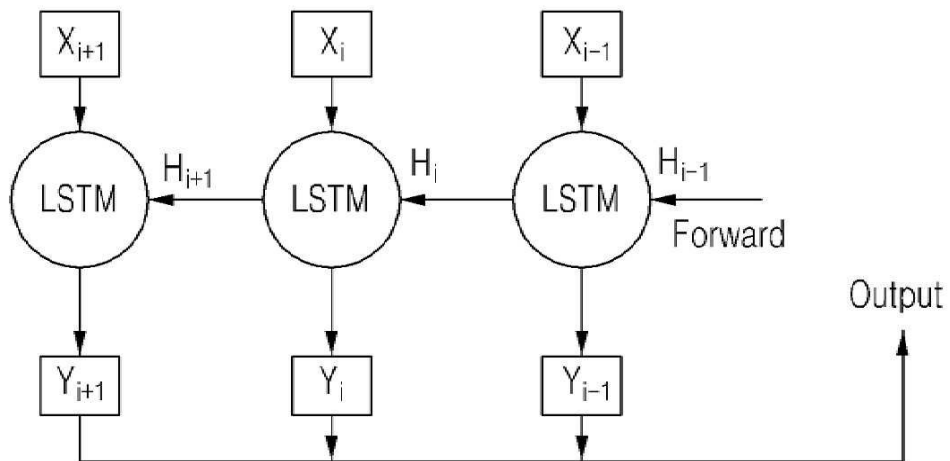
도면3



도면4



도면5



도면6

