

(19) 대한민국특허청(KR)
(12) 공개특허공보(A)(11) 공개번호 10-2025-0073758
(43) 공개일자 2025년05월27일

(51) 국제특허분류(Int. Cl.)

A61B 5/16 (2006.01) A61B 5/00 (2021.01)
A61B 5/01 (2021.01) A61B 5/024 (2025.01)
A61B 5/0531 (2021.01) G16H 10/20 (2018.01)
G16H 50/20 (2018.01)

(52) CPC특허분류

A61B 5/165 (2013.01)
A61B 5/01 (2021.01)

(21) 출원번호 10-2023-0160075

(22) 출원일자 2023년11월20일

심사청구일자 2023년11월20일

(71) 출원인

경희대학교 산학협력단

경기도 용인시 기흥구 덕영대로 1732 (서천동, 경희대학교 국제캠퍼스내)

(72) 발명자

이원희

경기도 용인시 기흥구 중부대로 184, B-334호(영덕동, 희스유타워)

김태성

경기도 수원시 팔달구 권광로 373, 106동 601호(우만동, 월드메르디앙)

(뒷면에 계속)

(74) 대리인

김연권

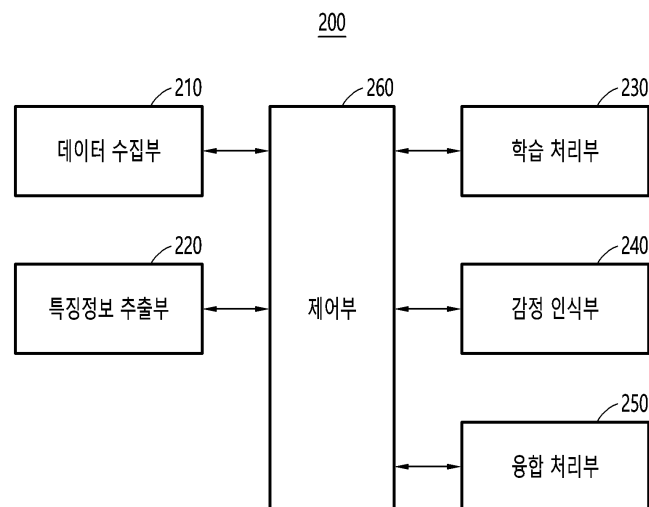
전체 청구항 수 : 총 19 항

(54) 발명의 명칭 멀티모달 데이터의 융합을 기반으로 감정을 인식하기 위한 장치 및 방법

(57) 요약

본 발명은 멀티 모달리티의 공통적 및 특이적 특징을 기반으로 감정을 인지하는 멀티모달 융합 학습 네트워크와 데이터 불균형 환경에서의 감정 인식을 위한 contrastive loss 함수를 개선하여 멀티모달 데이터의 융합을 기반으로 감정을 인식하기 위한 기술에 관한 것이다.

대표도 - 도2



(52) CPC특허분류

A61B 5/024 (2025.01)
A61B 5/0531 (2021.01)
A61B 5/4803 (2013.01)
A61B 5/7267 (2013.01)
G16H 10/20 (2021.08)
G16H 50/20 (2018.01)

조혜정

충청남도 천안시 서북구 월봉4로 88-11

(72) 발명자

김태성

인천광역시 서구 가정로 437, 111동 302호

한주혁

서울특별시 송파구 올림픽로4길 15, 10동 902호

이 발명을 지원한 국가연구개발사업

과제고유번호	1711193920
과제번호	2017-0-00655-007
부처명	과학기술정보통신부
과제관리(전문)기관명	정보통신기획평가원
연구사업명	(혁신법)디지털콘텐츠원천기술개발(R&D, 정보화)_2023
연구과제명	다양한 디지털 제품의 UX 평가를 위한 Lean UX 핵심기술 및 플랫폼(7/8)
기 여 율	50/100
과제수행기관명	경희대학교 산학협력단
연구기간	2017.04.01 ~ 2024.12.31

이 발명을 지원한 국가연구개발사업

과제고유번호	1711180606
과제번호	2021R1C1C1009436
부처명	과학기술정보통신부
과제관리(전문)기관명	한국연구재단
연구사업명	우수신진연구
연구과제명	멀티모달 뇌영상 데이터를 이용한 딥러닝 기반 조현병 환자의 뇌연령 및 가속 노화

예측 알고리즘 개발(3/3)

기 여 율	45/100
과제수행기관명	경희대학교 산학협력단
연구기간	2021.03.01 ~ 2024.02.29

이 발명을 지원한 국가연구개발사업

과제고유번호	1711179316
과제번호	00155911
부처명	과학기술정보통신부
과제관리(전문)기관명	정보통신기획평가원
연구사업명	(혁신법)인공지능융합혁신인재양성_2023
연구과제명	[2023 정부] 인공지능융합혁신인재양성(경희대학교) (2/4)
기 여 율	5/100
과제수행기관명	경희대학교 산학협력단
연구기간	2022.07.01 ~ 2025.12.31

명세서

청구범위

청구항 1

멀티모달 데이터의 융합을 기반으로 감정을 인식하는 장치에 있어서,

복수의 모달리티 데이터를 포함하는 멀티모달 데이터를 수집하는 데이터 수집부;

사전에 학습된 모델들을 이용하여, 상기 복수의 모달리티 데이터 각각으로부터 복수의 특징정보를 추출하되, 상기 각각의 모달리티 데이터 별로 고유 특징(specific feature) 및 상기 모달리티 데이터의 공통 특징(invariant feature)을 구분하여 추출하는 특징정보 추출부;

상기 추출된 복수의 특징정보를 학습하되, 상기 각각의 모달리티 데이터 별로 고유한 특징 및 상기 모달리티 데이터의 공통된 특징을 이용하여 학습하는 학습 처리부;

상기 학습된 특징을 융합하는 융합 처리부; 및

상기 융합된 데이터를 기반으로 감정을 인식하는 감정 인식부

를 포함하는 멀티모달 데이터 융합 기반의 감정 인식 장치.

청구항 2

제1항에 있어서,

상기 데이터 수집부는,

텍스트 데이터, 음성 데이터, 또는 생체신호 데이터 중에서 적어도 둘 이상의 모달리티 데이터를 수집하는 것을 특징으로 하는 멀티모달 데이터 융합 기반의 감정 인식 장치.

청구항 3

제2항에 있어서,

상기 생체신호 데이터는,

피부전도도, 손목 피부온도, 맥파 데이터에 대한 시간, 주파수, 시간-주파수 도메인별 특징을 포함하는 것을 특징으로 하는 멀티모달 데이터 융합 기반의 감정 인식 장치.

청구항 4

제1항에 있어서,

상기 융합 처리부는,

상기 모달리티 데이터에 대한 중요도를 기반으로 산출된 가중치를 적용하여 상기 학습된 특징을 융합하는 것을 특징으로 하는 멀티모달 데이터 융합 기반의 감정 인식 장치.

청구항 5

제4항에 있어서,

상기 융합 처리부는,

완전 연결 층(fully connected layer)으로 구성된 피드 포워드 층(feed forward layer)와 시그모이드(sigmoid) 함수를 적용하여 모달리티 데이터 별 중요도에 따라 0 내지 1사이의 가중치를 학습하고,

상기 학습된 가중치를 상기 고유 특징(specific feature)과 상기 공통 특징(invariant feature)에 대해 적용하는 것을 특징으로 하는 멀티모달 데이터 융합 기반의 감정 인식 장치.

청구항 6

제5항에 있어서,

상기 감정 인식부는,

상기 모달리티 데이터 마다 생성되는 고유 특징(specific feature)과 상기 공통 특징(invariant feature)이 감정을 인식하는데 얼마나 기여하는지를 반영해서 감정을 인식하는 것을 특징으로 하는 멀티모달 데이터 융합 기반의 감정 인식 장치.

청구항 7

제1항에 있어서,

상기 학습 처리부는,

서로 다른 모달리티 데이터의 고유 특징(specific feature)에 대해서는 거리가 멀어지도록 학습하고, 공통 특징(invariant feature)에 대해서는 거리가 가까워지도록 학습하며,

동일한 모달리티 데이터의 고유 특징(specific feature)과 공통 특징(invariant feature) 사이에 대해서는 거리가 멀어지도록 학습하는 것을 특징으로 하는 멀티모달 데이터 융합 기반의 감정 인식 장치.

청구항 8

제1항에 있어서,

상기 학습 처리부는,

동일한 감정 레이블을 가진 멀티모달 데이터에 대해서는 유사한 데이터 분포를 형성하도록 학습하고,

다른 감정 레이블을 가진 멀티모달 데이터에 대해서는 구별되는 데이터 분포를 형성하도록 학습하는 것을 특징으로 하는 멀티모달 데이터 융합 기반의 감정 인식 장치.

청구항 9

제8항에 있어서,

상기 학습 처리부는,

임의의 멀티모달 데이터를 앵커(anchor) 데이터로 지정하고,

상기 앵커 데이터와 같은 클래스를 갖는 멀티모달 데이터에 대해서는 포지티브 샘플로 지정하고,

상기 앵커 데이터와 다른 클래스를 갖는 멀티모달 데이터에 대해서는 네거티브 샘플로 지정하며,

상기 학습 처리부는,

대비 손실(contrastive loss)을 활용하여 상기 지정된 포지티브 샘플은 상기 앵커 데이터와 유사하도록 학습하고, 상기 지정된 네거티브 샘플은 상기 앵커 데이터와 구별되도록 학습하는 것을 특징으로 하는 멀티모달 데이터 융합 기반의 감정 인식 장치.

청구항 10

제9항에 있어서,

상기 학습 처리부는,

상기 대비 손실(contrastive loss)을 산출하는 과정에서,

동일한 멀티모달(intra-modal) 데이터 내에서 연산하는 유사도와, 서로 다른 멀티모달(inter modal) 데이터 간 연산하는 유사도를 각각 독립적으로 산출하되,

상기 동일한 멀티모달(intra-modal) 데이터 내에서는 고유 특징(specific feature)과 공통 특징(invariant feature)을 모두 적용하여 유사도를 연산하고,

서로 다른 멀티모달(inter modal) 데이터 간에는 공통 특징(invariant feature)만을 적용하여 유사도를 연산하는 멀티모달 데이터 융합 기반의 감정 인식 장치.

청구항 11

제9항에 있어서,

상기 학습 처리부는,

상기 지정된 포지티브 샘플과 상기 앵커 데이터의 유사도는 1이 되도록 하고, 상기 지정된 네거티브 샘플과 상기 앵커 데이터의 유사도는 -1이 되도록 학습하는 것을 특징으로 하는 멀티모달 데이터 융합 기반의 감정 인식 장치.

청구항 12

제9항에 있어서,

상기 학습 처리부는,

상기 임의의 멀티모달 데이터를 앵커(anchor) 데이터로 지정하되, 멀티모달 데이터의 감정을 기준으로 하여 순차적으로 지정되도록 상기 앵커(anchor) 데이터를 지정하는 것을 특징으로 하는 멀티모달 데이터 융합 기반의 감정 인식 장치.

청구항 13

멀티모달 데이터의 융합을 기반으로 감정을 인식하는 방법에 있어서,

복수의 모달리티 데이터를 포함하는 멀티모달 데이터를 수집하는 단계;

사전에 학습된 모델들을 이용하여, 상기 복수의 모달리티 데이터 각각으로부터 복수의 특징정보를 추출하되, 상기 각각의 모달리티 데이터 별로 고유 특징(specific feature) 및 상기 모달리티 데이터의 공통 특징(invariant feature)을 구분하여 추출하는 단계;

상기 추출된 복수의 특징정보를 학습하되, 상기 각각의 모달리티 데이터 별로 고유한 특징 및 상기 모달리티 데이터의 공통된 특징을 이용하여 학습하는 단계; 및

상기 학습된 특징을 융합하고, 상기 융합된 데이터를 기반으로 감정을 인식하는 단계를 포함하는 멀티모달 데이터 융합 기반의 감정 인식 방법.

청구항 14

제13항에 있어서,

상기 복수의 모달리티 데이터를 포함하는 멀티모달 데이터를 수집하는 단계는,

텍스트 데이터, 음성 데이터, 또는 생체신호 데이터 중에서 적어도 둘 이상의 모달리티 데이터를 수집하는 단계를 포함하되,

상기 생체신호 데이터는,

피부전도도, 손목 피부온도, 맥파 데이터에 대한 시간, 주파수, 시간-주파수 도메인별 특징을 포함하는 것을 특징으로 하는 멀티모달 데이터 융합 기반의 감정 인식 방법.

청구항 15

제13항에 있어서,

상기 학습된 특징을 융합하고, 상기 융합된 데이터를 기반으로 감정을 인식하는 단계는,

상기 모달리티 데이터에 대한 중요도를 기반으로 산출된 가중치를 적용하여 상기 학습된 특징을 융합하는 단계를 포함하는 것을 특징으로 하는 멀티모달 데이터 융합 기반의 감정 인식 방법.

청구항 16

제15항에 있어서,

상기 학습된 특징을 융합하고, 상기 융합된 데이터를 기반으로 감정을 인식하는 단계는,

완전 연결 층(fully connected layer)으로 구성된 피드 포워드 층(feed forward layer)와 시그모이드(sigmoid) 함수를 적용하여 모달리티 데이터 별 중요도에 따라 0 내지 1사이의 가중치를 학습하는 단계;

상기 학습된 가중치를 상기 고유 특징(specific feature)과 상기 공통 특징(invariant feature)에 대해 적용하는 단계; 및

상기 모달리티 데이터 마다 생성되는 고유 특징(specific feature)과 상기 공통 특징(invariant feature)이 감정을 인식하는데 얼마나 기여하는지를 반영해서 감정을 인식하는 단계

를 포함하는 것을 특징으로 하는 멀티모달 데이터 융합 기반의 감정 인식 방법.

청구항 17

제13항에 있어서,

상기 추출된 복수의 특징정보를 학습하는 단계는,

서로 다른 모달리티 데이터의 고유 특징(specific feature)에 대해서는 거리가 멀어지도록 학습하고, 공통 특징(invariant feature)에 대해서는 거리가 가까워지도록 학습하는 단계; 및

동일한 모달리티 데이터의 고유 특징(specific feature)과 공통 특징(invariant feature) 사이에 대해서는 거리가 멀어지도록 학습하는 단계

를 포함하는 것을 특징으로 하는 멀티모달 데이터 융합 기반의 감정 인식 방법.

청구항 18

제13항에 있어서,

상기 추출된 복수의 특징정보를 학습하는 단계는,

임의의 멀티모달 데이터를 앵커(anchor) 데이터로 지정하는 단계;

상기 앵커 데이터와 같은 클래스를 갖는 멀티모달 데이터에 대해서는 포지티브 샘플로 지정하고, 상기 앵커 데이터와 다른 클래스를 갖는 멀티모달 데이터에 대해서는 네거티브 샘플로 지정하는 단계

대비 손실(contrastive loss)을 활용하여 상기 지정된 포지티브 샘플은 상기 앵커 데이터와 유사하도록 학습하고, 상기 지정된 네거티브 샘플은 상기 앵커 데이터와 구별되도록 학습하는 단계

를 포함하되,

상기 지정된 포지티브 샘플과 상기 앵커 데이터의 유사도는 1이 되도록 하고, 상기 지정된 네거티브 샘플과 상기 앵커 데이터의 유사도는 -1이 되도록 학습하고,

상기 대비 손실(contrastive loss)을 활용하는 과정에서,

동일한 멀티모달(intra-modal) 데이터 내에서 연산하는 유사도와, 서로 다른 멀티모달(inter modal) 데이터 간 연산하는 유사도를 각각 독립적으로 산출하되,

상기 동일한 멀티모달(intra-modal) 데이터 내에서는 고유 특징(specific feature)과 공통 특징(invariant feature)을 모두 적용하여 유사도를 연산하고,

서로 다른 멀티모달(inter modal) 데이터 간에는 공통 특징(invariant feature)만을 적용하여 유사도를 연산하는 것을 특징으로 하는 멀티모달 데이터 융합 기반의 감정 인식 방법.

청구항 19

제18항에 있어서,

상기 임의의 멀티모달 데이터를 앵커(anchor) 데이터로 지정하는 단계는,

임의의 멀티모달 데이터를 앵커(anchor) 데이터로 지정하되, 멀티모달 데이터의 감정을 기준으로 하여 순차적으로 지정되도록 상기 앵커(anchor) 데이터를 지정하는 단계

를 포함하는 것을 특징으로 하는 멀티모달 데이터 융합 기반의 감정 인식 방법.

발명의 설명

기술 분야

[0001] 본 발명은 멀티 모달리티의 공통 특징 및 고유 특징을 기반으로 감정을 인지하는 멀티모달 융합 학습 네트워크와 데이터 불균형 환경에서의 감정 인식을 위한 대조 손실(contrastive loss) 함수를 개선하여 멀티모달 데이터의 융합을 기반으로 감정을 인식하기 위한 기술에 관한 것이다.

배경 기술

[0002] 감정은 일상적으로 마주하는 상황에 대한 반응을 조절하고, 주변 환경과의 상호작용을 가능하게 하는 중요한 역할을 한다. 인간의 감정을 인식하는 기술은 삶의 질 개선, 사용자 경험 평가, 인간-컴퓨터 상호작용(Human-Computer Interaction, HCI) 등 다양한 분야에서 활용될 수 있어 인공지능을 이용한 감정 인식 연구가 활발히 진행되고 있다.

[0003] 감정은 인체 전반에 걸쳐 여러 양상으로 나타난다. 멀티모달 데이터는 단일모달 데이터보다 더 다양한 정보를 제공하기 때문에 멀티모달 데이터를 활용한 감정 인식 연구들이 다수 존재한다.

[0004] 인공지능을 이용하여 감정을 인식하는 기술은 다양한 데이터 소스를 활용하여 사람들의 감정 상태를 이해하고 분석하는 기술을 의미한다.

[0005] 이를 위해, 자연어 처리 (NLP), 음성 인식, 이미지 및 영상 인식 (컴퓨터 비전), 감정 데이터셋과 머신러닝, 심층 학습 (Deep Learning), 시계열 분석, 재귀 신경망 (RNN) 및 장단기 메모리 (LSTM) 등의 기술들이 사용될 수 있다.

[0006] 자연어 처리 (NLP)는 텍스트 데이터를 처리하고 이해하는 데 사용되며, 감정 분석을 위해 감정 관련 단어나 문맥을 파악하고, 긍정적인지 부정적인지 판단할 수 있다.

[0007] 음성 인식은 음성 데이터를 분석하여 감정을 파악하는 기술로서, 음성 톤, 강도, 억양 등을 통해 감정을 추정할 수 있다.

[0008] 또한, 이미지 및 영상 인식 (컴퓨터 비전)은 얼굴 표정, 몸짓 등을 분석하여 감정을 인식하는 기술로서, 얼굴 인식 기술을 통해 눈, 입, 눈썹 등의 특징을 파악하고 감정을 판단할 수 있다.

[0009] 인공지능 모델은 대량의 감정 관련 데이터를 학습하여 감정을 인식할 수 있으며, 이 데이터는 레이블링되어 있어서 모델이 어떤 감정과 관련이 있는지 학습할 수 있다.

[0010] 또한, 심층 신경망을 활용하여 복잡한 데이터에서 패턴을 학습하고 감정을 인식할 수 있고, 음성이나 동영상 데이터의 경우, 시간에 따른 감정 변화를 파악하기 위해 시계열 분석 기술을 활용할 수 있으며, 시간에 따른 연속적인 데이터를 다루는 데 재귀 신경망 (RNN) 및 장단기 메모리 (LSTM)를 사용할 수 있다.

[0011] 모달리티별 데이터의 고유한 특성과 표현 방식으로 멀티모달 데이터 사이에 이질성이 존재한다. 이러한 이유로 멀티모달 데이터를 이용한 인공지능 모델 학습에 어려움이 있다. 따라서, 각 모달리티의 특이적 특성을 보존하면서도 공통된 차원에서 표현될 수 있도록 융합하는 과정은 멀티모달 인공지능 모델을 학습하는데 필수적이다.

[0012] 또한, 감정의 특성상 극단적인 감정에 대한 데이터 수집의 한계가 존재한다. 데이터 불균형 상황에서 데이터 수가 많은 일부 감정에 편향되어 감정을 예측하는 오류는 빈번히 발생한다.

선행기술문헌

특허문헌

[0013] (특허문헌 0001) 한국등록특허 제10-2275436호 "실시간 감정인식을 위한 영상 및 eeg 신호의 융합 기술"

(특허문헌 0002) 한국등록특허 제10-2243294호 "영상 및 eeg 신호 융합을 이용한 인간 감정상태 인식 방법 및 시스템"

(특허문헌 0003) 한국등록특허 제10-2570279호 "감정 인식을 위한 학습 방법, 감정을 인식하는 방법 및 장치"

발명의 내용

해결하려는 과제

- [0014] 본 발명은 모달리티별 상호 보완적인 특징을 추출하고 이들을 융합시키기 위한 네트워크를 설계하고, 기존의 대조 손실을 개선하여 불균형한 데이터 환경에서도 모든 감정에 대해 고루 예측할 수 있는 감정 인식 모델을 제시하는 것을 목적으로 한다.
- [0015] 본 발명은 모달리티 데이터 특성에 맞게 추출된 특징들은 공통된 특징 공간에 위치시키는 모달리티 공통(modality-invariant) 특징과 모달리티별로 고유한 특성을 나타내는 모달리티 고유(modality-specific) 특징을 생성하는 트랜스포머 기반의 모달리티 표현 인코더(Modality Representations Encoder)를 통과한다. 추출된 특징들은 Gated Fusion Block을 거쳐 융합되며 이를 기반으로 감정을 예측하는 것을 목적으로 한다.
- [0016] 본 발명은 특징분포가 수렴하지 못해 모델 학습에 방해되는 문제를 방지하며, 분별력 있는 특징분포의 원활한 학습을 가능하게 하는 것을 목적으로 한다.
- [0017] 본 발명은 샘플 수가 많은 감정에 편향되어 학습되는 것을 방지하고 샘플 수가 적은 감정에 대해서도 준수한 성능을 제공하는 것을 목적으로 한다.
- [0018] 본 발명은 다양하고 범용적인 모달리티 데이터 사용과 세밀한 감정 인식을 가능하게 하며, 특히 제한적인 데이터 상황에서 활용 가능성이 높은 감정 인식 기술을 제공하는 것을 목적으로 한다.

과제의 해결 수단

- [0019] 일실시예에 따른 멀티모달 데이터 융합 기반의 감정 인식 장치는 복수의 모달리티 데이터를 포함하는 멀티모달 데이터를 수집하는 데이터 수집부, 사전에 학습된 모델들을 이용하여, 상기 복수의 모달리티 데이터 각각으로부터 복수의 특징정보를 추출하되, 상기 각각의 모달리티 데이터 별로 고유 특징(specific feature) 및 상기 모달리티 데이터의 공통 특징(invariant feature)을 구분하여 추출하는 특징정보 추출부, 상기 추출된 복수의 특징정보를 학습하되, 상기 각각의 모달리티 데이터 별로 고유한 특징 및 상기 모달리티 데이터의 공통된 특징을 이용하여 학습하는 학습 처리부, 상기 학습된 특징을 융합하는 융합 처리부, 및 상기 융합된 데이터를 기반으로 감정을 인식하는 감정 인식부를 포함할 수 있다.
- [0020] 일실시예에 따른 상기 데이터 수집부는, 텍스트 데이터, 음성 데이터, 또는 생체신호 데이터 중에서 적어도 둘 이상의 모달리티 데이터를 수집할 수 있다.
- [0021] 일실시예에 따른 상기 생체신호 데이터는, 피부전도도, 손목 피부온도, 맥파 데이터에 대한 시간, 주파수, 시간-주파수 도메인별 특징을 포함할 수 있다.
- [0022] 일실시예에 따른 상기 융합 처리부는, 상기 모달리티 데이터에 대한 중요도를 기반으로 산출된 가중치를 적용하여 상기 학습된 특징을 융합할 수 있다.
- [0023] 일실시예에 따른 상기 융합 처리부는, 완전 연결 층(fully connected layer)으로 구성된 피드 포워드 층(feed forward layer)와 시그모이드(sigmoid) 함수를 적용하여 모달리티 데이터 별 중요도에 따라 0 내지 1사이의 가중치를 학습하고, 상기 학습된 가중치를 상기 고유 특징(specific feature)과 상기 공통 특징(invariant feature)에 대해 적용할 수 있다.
- [0024] 일실시예에 따른 상기 감정 인식부는, 모달리티 데이터 마다 생성되는 고유 특징(specific feature)과 상기 공통 특징(invariant feature)이 감정을 인식하는데 얼마나 기여하는지를 반영해서 감정을 인식할 수 있다.
- [0025] 일실시예에 따른 상기 학습 처리부는, 서로 다른 모달리티 데이터의 고유 특징(specific feature)에 대해서는 거리가 멀어지도록 학습하고, 공통 특징(invariant feature)에 대해서는 거리가 가까워지도록 학습할 수 있다. 동일한 모달리티 데이터의 고유 특징(specific feature)과 공통 특징(invariant feature) 사이에 대해서는 거

리가 멀어지도록 학습할 수 있다.

- [0026] 일실시예에 따른 상기 학습 처리부는, 동일한 감정 레이블을 가진 멀티모달 데이터에 대해서는 유사한 데이터 분포를 형성하도록 학습하고, 다른 감정 레이블을 가진 멀티모달 데이터에 대해서는 구별되는 데이터 분포를 형성하도록 학습할 수 있다.
- [0027] 일실시예에 따른 상기 학습 처리부는, 임의의 멀티모달 데이터를 앵커(anchor) 데이터로 지정하고, 상기 앵커 데이터와 같은 클래스를 갖는 멀티모달 데이터에 대해서는 포지티브 샘플로 지정하고, 상기 앵커 데이터와 다른 클래스를 갖는 멀티모달 데이터에 대해서는 네거티브 샘플로 지정하며, 상기 학습 처리부는, 대비 손실(contrastive loss)을 활용하여 상기 지정된 포지티브 샘플은 상기 앵커 데이터와 유사하도록 학습하고, 상기 지정된 네거티브 샘플은 상기 앵커 데이터와 구별되도록 학습할 수 있다.
- [0028] 일실시예에 따른 상기 학습 처리부는, 상기 대비 손실(contrastive loss)을 산출하는 과정에서, 동일한 멀티모달(intra-modal) 데이터 내에서 연산하는 유사도와, 서로 다른 멀티모달(inter modal) 데이터 간 연산하는 유사도를 각각 독립적으로 산출하되, 상기 동일한 멀티모달(intra-modal) 데이터 내에서는 고유 특징(specific feature)과 공통 특징(invariant feature)을 모두 적용하여 유사도를 연산하고, 서로 다른 멀티모달(inter modal) 데이터 간에는 공통 특징(invariant feature)만을 적용하여 유사도를 연산할 수 있다.
- [0029] 일실시예에 따른 상기 학습 처리부는, 상기 지정된 포지티브 샘플과 상기 앵커 데이터의 유사도는 1이 되도록 하고, 상기 지정된 네거티브 샘플과 상기 앵커 데이터의 유사도는 -1이 되도록 학습할 수 있다.
- [0030] 일실시예에 따른 상기 학습 처리부는, 상기 임의의 멀티모달 데이터를 앵커(anchor) 데이터로 지정하되, 멀티모달 데이터의 감정을 기준으로 하여 순차적으로 지정되도록 상기 앵커(anchor) 데이터를 지정할 수 있다.
- [0031] 일실시예에 따른 멀티모달 데이터 융합 기반의 감정 인식 방법은 복수의 모달리티 데이터를 포함하는 멀티모달 데이터를 수집하는 단계, 사전에 학습된 모델들을 이용하여, 상기 복수의 모달리티 데이터 각각으로부터 복수의 특징정보를 추출하되, 상기 각각의 모달리티 데이터 별로 고유 특징(specific feature) 및 상기 모달리티 데이터의 공통 특징(invariant feature)을 구분하여 추출하는 단계, 상기 추출된 복수의 특징정보를 학습하되, 상기 각각의 모달리티 데이터 별로 고유한 특징 및 상기 모달리티 데이터의 공통된 특징을 이용하여 학습하는 단계, 및 상기 학습된 특징을 융합하고, 상기 융합된 데이터를 기반으로 감정을 인식하는 단계를 포함할 수 있다.
- [0032] 일실시예에 따른 상기 복수의 모달리티 데이터를 포함하는 멀티모달 데이터를 수집하는 단계는, 텍스트 데이터, 음성 데이터, 또는 생체신호 데이터 중에서 적어도 둘 이상의 모달리티 데이터를 수집하는 단계를 포함하되, 상기 생체신호 데이터는, 피부전도도, 손목 피부온도, 맥파 데이터에 대한 시간, 주파수, 시간-주파수 도메인별 특징을 포함한다.
- [0033] 일실시예에 따른 상기 학습된 특징을 융합하고, 상기 융합된 데이터를 기반으로 감정을 인식하는 단계는, 상기 모달리티 데이터에 대한 중요도를 기반으로 산출된 가중치를 적용하여 상기 학습된 특징을 융합하는 단계를 포함할 수 있다.
- [0034] 일실시예에 따른 상기 학습된 특징을 융합하고, 상기 융합된 데이터를 기반으로 감정을 인식하는 단계는, 완전 연결 층(fully connected layer)으로 구성된 피드 포워드 층(feed forward layer)와 시그모이드(sigmoid) 함수를 적용하여 모달리티 데이터 별 중요도에 따라 0 내지 1사이의 가중치를 학습하는 단계, 상기 학습된 가중치를 상기 고유 특징(specific feature)과 상기 공통 특징(invariant feature)에 대해 적용하는 단계, 및 상기 모달리티 데이터 마다 생성되는 고유 특징(specific feature)과 상기 공통 특징(invariant feature)이 감정을 인식하는데 얼마나 기여하는지를 반영해서 감정을 인식하는 단계를 포함할 수 있다.
- [0035] 일실시예에 따른 상기 추출된 복수의 특징정보를 학습하는 단계는, 서로 다른 모달리티 데이터의 고유 특징(specific feature)에 대해서는 거리가 멀어지도록 학습하고, 공통 특징(invariant feature)에 대해서는 거리가 가까워지도록 학습하는 단계, 및 동일한 모달리티 데이터의 고유 특징(specific feature)과 공통 특징(invariant feature) 사이에 대해서는 거리가 멀어지도록 학습하는 단계를 포함할 수 있다.
- [0036] 일실시예에 따른 상기 추출된 복수의 특징정보를 학습하는 단계는, 임의의 멀티모달 데이터를 앵커(anchor) 데이터로 지정하는 단계, 상기 앵커 데이터와 같은 클래스를 갖는 멀티모달 데이터에 대해서는 포지티브 샘플로 지정하고, 상기 앵커 데이터와 다른 클래스를 갖는 멀티모달 데이터에 대해서는 네거티브 샘플로 지정하는 단계, 대비 손실(contrastive loss)을 활용하여 상기 지정된 포지티브 샘플은 상기 앵커 데이터와 유사하도록 학습하고, 상기 지정된 네거티브 샘플은 상기 앵커 데이터와 구별되도록 학습하는 단계를 포함하되, 상기 지정

된 포지티브 샘플과 상기 앵커 데이터의 유사도는 1이 되도록 하고, 상기 지정된 네거티브 샘플과 상기 앵커 데이터의 유사도는 -1이 되도록 학습하고, 상기 대비 손실(contrastive loss)을 활용하는 과정에서, 동일한 멀티모달(intra-modal) 데이터 내에서 연산하는 유사도와, 서로 다른 멀티모달(inter modal) 데이터 간 연산하는 유사도를 각각 독립적으로 산출하되, 상기 동일한 멀티모달(intra-modal) 데이터 내에서는 고유 특징(specific feature)과 공통 특징(invariant feature)을 모두 적용하여 유사도를 연산하고, 서로 다른 멀티모달(inter modal) 데이터 간에는 공통 특징(invariant feature)만을 적용하여 유사도를 연산할 수 있다.

[0037] 일실시예에 따른 상기 임의의 멀티모달 데이터를 앵커(anchor) 데이터로 지정하는 단계는, 임의의 멀티모달 데이터를 앵커(anchor) 데이터로 지정하되, 멀티모달 데이터의 감정을 기준으로 하여 순차적으로 지정되도록 상기 앵커(anchor) 데이터를 지정하는 단계를 포함할 수 있다.

발명의 효과

[0038] 일실시예에 따르면, 모달리티별 상호 보완적인 특징을 추출하고 이들을 융합시키기 위한 네트워크를 설계하고, 기존의 contrastive loss를 개선하여 불균형한 데이터 환경에서도 모든 감정에 대해 고루 예측할 수 있는 감정 인식 모델을 제시할 수 있다.

[0039] 일실시예에 따르면, 모달리티 데이터 특성에 맞게 추출된 특징들은 공통된 특징 공간에 위치시키는 모달리티 공통(modality-invariant) 특징과 모달리티별로 고유한 특성을 나타내는 모달리티 고유(modality-specific) 특징을 생성하는 트랜스포머 기반의 모달리티 표현 인코더를 통과한다. 추출된 특징들은 Gated Fusion Block을 거쳐 융합되며 이를 기반으로 감정을 예측할 수 있다.

[0040] 일실시예에 따르면, 특징분포가 수렴하지 못해 모델 학습에 방해되는 문제를 방지하며, 분별력 있는 특징분포의 원활한 학습을 가능하게 할 수 있다.

[0041] 일실시예에 따르면, 샘플 수가 많은 감정에 편향되어 학습되는 것을 방지하고 샘플 수가 적은 감정에 대해서도 준수한 성능을 제공할 수 있다.

[0042] 일실시예에 따르면, 다양하고 범용적인 모달리티 데이터 사용과 세밀한 감정 인식을 가능하게 하며, 특히 제한적인 데이터 상황에서 활용 가능성이 높은 감정 인식 기술을 제공할 수 있다.

도면의 간단한 설명

[0043] 도 1은 감정 인식을 위한 멀티모달 융합 학습 네트워크를 도식화한 도면이다.

도 2는 일실시예에 따른 멀티모달 데이터 융합 기반의 감정 인식 장치를 설명하는 도면이다.

도 3은 있도록 트랜스포머 모델의 인코더로 구성된 modality specific encoder를 통해 각 모달리티마다 modality specific 특징을 생성하는 실시예를 설명하는 도면이다.

도 4는 모달리티의 중요도에 따라 가중하여 이들을 융합하는 Gated Fusion Block을 설명하는 도면이다.

도 5a는 각 Sample 내 특징 사이에서 계산하는 위딘 샘플 로스(Within-sample Loss)를 설명하는 도면이다.

도 5b는 서로 다른 두 샘플 사이의 특징끼리 계산하는 비트윈 샘플 로스(Between-sample Loss)를 설명하는 도면이다.

도 6은 위딘 샘플 로스(Within-sample Loss)에서 모달리티-고유(Modality-specific)와 모달리티-공통(Modality-invariant)을 대비하여 설명하는 도면이다.

도 7은 위딘 샘플 로스(Within-sample Loss)의 SOC loss와 CMD loss를 설명하는 도면이다.

도 8은 동일한 감정을 갖는 샘플들의 특징은 유사하도록, 다른 감정을 갖는 샘플들의 특징들은 차이가 나도록 하는 계산하는 비트윈 샘플 로스(Between-sample Loss)를 설명하는 도면이다.

도 9는 배치단위로 계산되는 계산하는 비트윈 샘플 로스(Between-sample Loss)를 설명하는 도면이다.

도 10은 본 발명에 따른 대조 손실로서, 같은 모달리티 특징 간 계산과 다른 모달리티 특징 간 계산을 구분하여 처리하는 계산하는 비트윈 샘플 로스(Between-sample Loss)를 설명하는 도면이다.

도 11은 감정에 따라 순차적으로 앵커가 선정되도록 하는 계산하는 비트윈 샘플 로스(Between-sample Loss)를

설명하는 도면이다.

도 12는 감정 인식을 위한 멀티모달 융합 학습 네트워크를 설명하는 도면이다.

도 13은 일실시예에 따른 멀티모달 데이터 융합 기반의 감정 인식 방법을 설명하는 도면이다.

발명을 실시하기 위한 구체적인 내용

- [0044] 본 명세서에 개시되어 있는 본 발명의 개념에 따른 실시예들에 대해서 특정한 구조적 또는 기능적 설명들은 단지 본 발명의 개념에 따른 실시예들을 설명하기 위한 목적으로 예시된 것으로서, 본 발명의 개념에 따른 실시예들은 다양한 형태로 실시될 수 있으며 본 명세서에 설명된 실시예들에 한정되지 않는다.
- [0045] 본 발명의 개념에 따른 실시예들은 다양한 변경들을 가할 수 있고 여러 가지 형태들을 가질 수 있으므로 실시예들을 도면에 예시하고 본 명세서에 상세하게 설명하고자 한다. 그러나, 이는 본 발명의 개념에 따른 실시예들을 특정한 개시형태들에 대해 한정하려는 것이 아니며, 본 발명의 사상 및 기술 범위에 포함되는 변경, 균등물, 또는 대체물을 포함한다.
- [0046] 제1 또는 제2 등의 용어를 다양한 구성요소들을 설명하는데 사용될 수 있지만, 상기 구성요소들은 상기 용어들에 의해 한정되어서는 안 된다. 상기 용어들은 하나의 구성요소를 다른 구성요소로부터 구별하는 목적으로만, 예를 들어 본 발명의 개념에 따른 권리 범위로부터 이탈되지 않은 채, 제1 구성요소는 제2 구성요소로 명명될 수 있고, 유사하게 제2 구성요소는 제1 구성요소로도 명명될 수 있다.
- [0047] 어떤 구성요소가 다른 구성요소에 "연결되어" 있다거나 "접속되어" 있다고 언급된 때에는, 그 다른 구성요소에 직접적으로 연결되어 있거나 또는 접속되어 있을 수도 있지만, 중간에 다른 구성요소가 존재할 수도 있다고 이해되어야 할 것이다. 반면에, 어떤 구성요소가 다른 구성요소에 "직접 연결되어" 있다거나 "직접 접속되어" 있다고 언급된 때에는, 중간에 다른 구성요소가 존재하지 않는 것으로 이해되어야 할 것이다. 구성요소들 간의 관계를 설명하는 표현들, 예를 들어 "~사이에"와 "바로~사이에" 또는 "~에 직접 이웃하는" 등도 마찬가지로 해석되어야 한다.
- [0048] 본 명세서에서 사용한 용어는 단지 특정한 실시예들을 설명하기 위해 사용된 것으로, 본 발명을 한정하려는 의도가 아니다. 단수의 표현은 문맥상 명백하게 다르게 뜻하지 않는 한, 복수의 표현을 포함한다. 본 명세서에서, "포함하다" 또는 "가지다" 등의 용어는 실시된 특징, 숫자, 단계, 동작, 구성요소, 부분품 또는 이들을 조합한 것이 존재함으로 지정하려는 것이지, 하나 또는 그 이상의 다른 특징들이나 숫자, 단계, 동작, 구성요소, 부분품 또는 이들을 조합한 것들의 존재 또는 부가 가능성을 미리 배제하지 않는 것으로 이해되어야 한다.
- [0049] 다르게 정의되지 않는 한, 기술적이거나 과학적인 용어를 포함해서 여기서 사용되는 모든 용어들은 본 발명이 속하는 기술 분야에서 통상의 지식을 가진 자에 의해 일반적으로 이해되는 것과 동일한 의미를 가진다. 일반적으로 사용되는 사전에 정의되어 있는 것과 같은 용어들은 관련 기술의 문맥상 가지는 의미와 일치하는 의미를 갖는 것으로 해석되어야 하며, 본 명세서에서 명백하게 정의하지 않는 한, 이상적이거나 과도하게 형식적인 의미로 해석되지 않는다.
- [0051] 이하, 실시예들을 첨부된 도면을 참조하여 상세하게 설명한다. 그러나, 특허출원의 범위가 이러한 실시예들에 의해 제한되거나 한정되는 것은 아니다. 각 도면에 제시된 동일한 참조 부호는 동일한 부재를 나타낸다.
- [0053] 도 1은 감정 인식을 위한 멀티모달 융합 학습 네트워크를 도식화한 도면(100)이다.
- [0054] 멀티모달 데이터(110)는 싱글모달 데이터 보다 더 다양한 정보를 제공할 수 있다. 그러나, 모달리티별 고유한 특성과 표현방식으로 멀티모달 데이터(110) 간 이질성 문제 해결 필요하다.
- [0055] 멀티모달 데이터(110)를 융합(Fusion)할 수 있는 기술이 중요한데, 본 발명에 따른 멀티모달 데이터 융합 기반의 감정 인식 장치는 텍스트 데이터, 음성 데이터, 또는 생체신호 데이터의 멀티모달 데이터(110)를 수집하고 이를 가공하여 융합할 수 있다.
- [0056] 멀티모달 데이터 융합 기반의 감정 인식 장치는 멀티모달 데이터(110)를 융합하는데, 각 모달리티 데이터의 잠재적인 정보를 활용하기 위해, 도면부호 120과 같이 각 모달리티 데이터로부터 사전학습된 모델이나 임의로 계

산한 특징들(130)을 추출한다.

- [0057] 또한, 트랜스포머 인코더(140)를 통해 모달리티 별 특화된 특징들과 모달리티의 공통적으로 드러나는 특징들을 학습할 수 있다.
- [0058] 트랜스포머 인코더(140)의 출력은 모달리티 데이터 별로 고유 특징(specific feature) 및 상기 모달리티 데이터의 공통 특징(invariant feature)에 해당한다.
- [0059] 이후 이러한 특징들을 잘 융합시키기 위해 gated fusion block(160)을 통과 하여 최종적으로 감정을 인식하게 된다.
- [0060] 결국, 모달리티별 상호 보완적인 특징을 추출하고 이들을 융합시키기 위한 네트워크를 설계하고, 기존의 contrastive loss를 개선하여 불균형한 데이터 환경에서도 모든 감정에 대해 고루 예측할 수 있는 감정 인식 모델을 제시할 수 있다.
- [0061] 또한, 본 발명에서는 모달리티 데이터 특성에 맞게 추출된 특징들은 공통된 특징 공간에 위치시키는 modality-invariant 특징과 모달리티별로 고유한 특성을 나타내는 modality-specific 특징을 생성하는 transformer 기반의 Modality Representations Encoder를 통과한다. 추출된 특징들은 Gated Fusion Block을 거쳐 융합되며 이를 기반으로 감정을 예측할 수 있다.
- [0062] 도 2는 일실시예에 따른 멀티모달 데이터 융합 기반의 감정 인식 장치를 설명하는 도면이다.
- [0063] 본 발명은 모달리티별 상호 보완적인 특징을 추출하고 이들을 융합시키기 위한 네트워크를 설계하고, 기존의 contrastive loss를 개선하여 불균형한 데이터 환경에서도 모든 감정에 대해 고루 예측할 수 있는 감정 인식 모델을 제시할 수 있다.
- [0064] 이를 위해, 멀티모달 데이터 융합 기반의 감정 인식 장치(200)는 데이터 수집부(210), 특징정보 추출부(220), 학습 처리부(230), 감정 인식부(240), 융합 처리부(250), 제어부(260)를 포함할 수 있다.
- [0065] 데이터 수집부(210)는 복수의 모달리티 데이터를 포함하는 멀티모달 데이터를 수집할 수 있다.
- [0066] 일례로, 데이터 수집부(210)는 텍스트 데이터, 음성 데이터, 또는 생체신호 데이터 중에서 적어도 둘 이상의 모달리티 데이터를 수집할 수 있다.
- [0067] 특히, 생체신호 데이터는 피부전도도, 손목 피부온도, 맥파 데이터에 대한 시간, 주파수, 시간-주파수 도메인별 특징을 포함할 수 있다.
- [0068] 특징정보 추출부(220)는 사전에 학습된 모델들을 이용하여, 상기 복수의 모달리티 데이터 각각으로부터 복수의 특징정보를 추출하되, 상기 각각의 모달리티 데이터 별로 고유 특징(specific feature) 및 상기 모달리티 데이터의 공통 특징(invariant feature)을 구분하여 추출할 수 있다.
- [0069] 특징정보 추출부(220)는 사전에 학습된 모델을 활용하여 다양한 모달리티 데이터에서 의미 있는 특징을 도출한다. 예를 들어, 이 모델들은 각 모달리티의 특성을 잘 파악하고 있는데, 예를 들어, 이미지 데이터의 경우 CNN(Convolutional Neural Network)을, 텍스트 데이터의 경우 LSTM(Long Short-Term Memory)을 활용할 수 있다. 이러한 사전 학습 모델을 통해 다양한 특징을 추출할 수 있다.
- [0070] 추출된 특징은 각 모달리티 데이터에 따라 고유한 특징과 해당 모달리티 데이터 간에 공통적으로 나타나는 특징으로 나뉜다. 예를 들어, 특정 이미지의 색상이나 형태는 해당 이미지의 특유한 특징일 수 있고, 여러 모달리티에서 나타나는 공통된 주제나 감정, 공통 변수들을 공통 특징으로 간주될 수 있다.
- [0071] 학습 처리부(230)는 추출된 복수의 특징정보를 학습하되, 상기 각각의 모달리티 데이터 별로 고유한 특징 및 상기 모달리티 데이터의 공통된 특징을 이용하여 학습할 수 있다.
- [0072] 학습 처리부(230)는 특징정보 추출부(220)에서 추출된 다양한 특징 정보를 학습하는 핵심적인 부분으로, 각 모달리티 데이터의 고유한 특징과 공통된 특징을 활용하여 학습할 수 있다.
- [0073] 먼저, 각 모달리티 데이터의 고유한 특징은 해당 모달리티가 가진 독특한 특성을 반영할 수 있다. 이러한 특징은 특정 모달리티의 특이성을 학습 처리부(230)에 제공하며, 예를 들어 이미지 데이터의 색상 패턴이나 텍스트 데이터의 문체 스타일과 같은 고유 특징을 이용하여 학습을 진행할 수 있다.
- [0074] 동시에, 각 모달리티 데이터의 공통된 특징은 서로 다른 모달리티 간에 공유되는 특성을 나타내는데, 이러한 공

통 특징은 학습 처리부(230)가 다양한 모달리티 간의 상호 작용을 이해하고, 이를 통합적으로 활용할 수 있도록 한다.

- [0075] 예를 들어, 여러 모달리티에서 나타나는 유사한 주제나 감정을 공통 특징으로 활용하여 시스템이 다양한 데이터로부터 통합된 정보를 추론하고 학습할 수 있다.
- [0076] 이러한 학습에 따라, 모델이 복잡한 데이터의 패턴을 이해하고 예측할 수 있다.
- [0077] 일례로, 학습 처리부(230)는 서로 다른 모달리티 데이터의 고유 특징(specific feature)에 대해서는 거리가 멀어지도록 학습하고, 공통 특징(invariant feature)에 대해서는 거리가 가까워지도록 학습할 수 있다. 동일한 모달리티 데이터의 고유 특징(specific feature)과 공통 특징(invariant feature) 사이에 대해서는 거리가 멀어지도록 학습할 수 있다.
- [0078] 뿐만 아니라, 학습 처리부(230)는 동일한 감정 레이블을 가진 멀티모달 데이터에 대해서는 유사한 데이터 분포를 형성하도록 학습하고, 다른 감정 레이블을 가진 멀티모달 데이터에 대해서는 구별되는 데이터 분포를 형성하도록 학습할 수 있다.
- [0079] 학습 처리부(230)는 특징들 간의 상호 관계를 효과적으로 학습함으로써 멀티모달 데이터 처리의 세분화된 측면을 강조할 수 있다.
- [0080] 먼저, 학습 처리부(230)는 하나의 샘플(멀티모달 데이터)에서 생성되는 6가지 특징들 사이에서의 연산을 이용하거나, 또는 서로 다른 두 샘플(멀티모달 데이터)에서 생성되는 특징들 사이에서 계산될 수 있다.
- [0081] 학습 처리부(230)는 동일한 모달리티 데이터의 고유 특징(specific feature)과 공통 특징(invariant feature) 사이에 대해서는 거리가 멀어지도록 학습할 수 있는데, 이는 모달리티의 고유한 특성과 공통적인 특성의 차이를 강조하며, 각 모달리티의 특성을 부각시킬 수 있다.
- [0082] 다음으로, 학습 처리부(230)는 서로 다른 모달리티 데이터의 고유 특징(specific feature)에 대해서는 거리가 멀어지도록 학습하고, 공통 특징(invariant feature)에 대해서는 거리가 가까워지도록 학습할 수 있다. 이는 서로 다른 멀티 모달 데이터 간의 독특한 특성을 부각시킬 수 있으며, 상호 작용하는 모든 모달리티 간에 공통적으로 드러나는 특징이 유사한 패턴을 형성하도록 한다. 이는 시스템이 다양한 상황에 대응할 수 있도록 데이터 간의 다양성을 학습하는 데 기여할 수 있다.
- [0083] 뿐만 아니라, 학습 처리부(230)는 동일한 감정 레이블을 가진 멀티모달 데이터에 대해서는 유사한 데이터 분포를 형성하도록 학습할 수 있다.
- [0084] 이는 감정과 관련된 멀티모달 데이터 간에 공통된 특성을 강조하며, 해당 감정 레이블에 따른 일관성 있는 특징을 추출할 수 있도록 한다.
- [0085] 또한, 학습 처리부(230)는 다른 감정 레이블을 가진 멀티모달 데이터에 대해서는 구별되는 데이터 분포를 형성하도록 학습할 수 있다. 이는 감정 레이블 간의 차이를 강조하며, 다양한 감정 상태에 대한 모델의 민감성을 높일 수 있다. 본 발명은 이러한 학습 방식에 따라 시스템이 감정을 효과적으로 구분하고 이해할 수 있다.
- [0086] 학습 처리부(230)는 임의의 멀티모달 데이터를 앵커(anchor) 데이터로 지정할 수 있다. 또한, 학습 처리부(230)는 앵커 데이터와 같은 클래스를 갖는 멀티모달 데이터에 대해서는 포지티브 샘플로 지정하고, 앵커 데이터와 다른 클래스를 갖는 멀티모달 데이터에 대해서는 네거티브 샘플로 지정할 수 있다.
- [0087] 일실시예에 따른 학습 처리부(230)는 학습의 효과를 극대화하기 위해 앵커(anchor) 데이터를 활용하는 고급 학습 방법을 제공할 수 있다.
- [0088] 앵커 기반 학습은 데이터 간의 상대적인 유사성을 명확하게 정의함으로써 모델이 데이터의 특성을 더욱 정교하게 파악할 수 있도록 한다.
- [0089] 먼저, 일실시예에 따른 학습 처리부(230)는 임의의 멀티모달 데이터를 앵커 데이터로 지정할 수 있다. 앵커 데이터는 모델이 학습하는 기준이 되는 중요한 데이터로서, 모델은 앵커를 중심으로 다른 데이터와의 상대적인 거리 및 관계를 학습할 수 있다.
- [0090] 앵커 데이터와 같은 클래스를 갖는 멀티모달 데이터에 대해서는 포지티브 샘플로 지정될 수 있는데, 이는 앵커 데이터와 같은 클래스에 속하는 데이터로서, 모델은 이러한 데이터들 간의 유사성을 강조하고 동일한 클래스에 대한 특징을 강화할 수 있다.

- [0091] 앵커 데이터와 다른 클래스를 갖는 멀티모달 데이터에 대해서는 네거티브 샘플로 지정될 수 있다. 이는 앵커와 다른 클래스에 속하는 데이터 간의 차이를 부각시킴으로써, 모델이 서로 다른 클래스 간의 경계를 명확하게 학습하도록 한다.
- [0092] 이러한 앵커 기반 학습은 모델이 데이터 간의 상대적인 유사성을 명확하게 인식하고, 특히 감정이나 카테고리과 같은 클래스에 따른 유사성을 더욱 강조하여 모델의 학습 성능을 향상시킬 수 있다.
- [0093] 학습 처리부(230)는 포지티브 샘플과 네거티브 샘플을 지정하게 되면, 대비 손실(contrastive loss)을 활용하여 지정된 포지티브 샘플은 상기 앵커 데이터와 유사하도록 학습하고, 지정된 네거티브 샘플은 앵커 데이터와 구별되도록 학습할 수 있다.
- [0094] 지정된 포지티브 샘플은 앵커 데이터와 동일한 클래스에 속하는 데이터로 지정될 수 있고, 이러한 샘플은 대비 손실을 통해 앵커 데이터와 유사하도록 학습될 수 있다. 즉, 모델은 앵커와 포지티브 샘플 간의 거리를 최소화하고, 두 데이터 간의 유사성을 강조하는 방향으로 학습된다.
- [0095] 지정된 네거티브 샘플은 앵커 데이터와 다른 클래스에 속하는 데이터로 지정되는데, 이 샘플은 대비 손실을 통해 앵커 데이터와 구별되도록 학습될 수 있다.
- [0096] 이러한 모델은 앵커와 네거티브 샘플 간의 거리를 최대화하고, 두 데이터 간의 차이를 강조하여 다른 클래스 간의 경계를 명확하게 학습하는 방향으로 업데이트될 수 있다.
- [0097] 이렇게 포지티브와 네거티브 샘플을 활용한 대비 손실 학습은 모델이 데이터 간의 유사성과 차이를 구분하면서 특징을 추출하도록 유도할 수 있다. 특히, 클래스 간의 구별력을 향상시켜 모델이 정확한 예측 및 분류를 수행할 수 있도록 한다.
- [0098] 일실시예에 따른 학습 처리부(230)는 대비 손실(contrastive loss)을 산출하는 과정에서, 동일한 멀티모달(intra-modal) 데이터 내에서 연산하는 유사도와, 서로 다른 멀티모달(inter modal) 데이터 간 연산하는 유사도를 각각 독립적으로 산출할 수 있다.
- [0099] 또한, 이 과정에서, 동일한 멀티모달(intra-modal) 데이터 내에서는 고유 특징(specific feature)과 공통 특징(invariant feature)을 모두 적용하여 유사도를 연산하고, 서로 다른 멀티모달(inter modal) 데이터 간에는 공통 특징(invariant feature)만을 적용하여 유사도를 연산할 수 있다.
- [0100] 일반적으로 대비 손실(Contrastive Loss)을 산출하는 과정에서, 학습 처리부(230)는 앵커 데이터, 포지티브 샘플, 네거티브 샘플 간의 유사도를 평가하는 방식으로 모델을 학습시킬 수 있다.
- [0101] 기존 대비 손실(Contrastive Loss)을 산출하는 과정에서의 분자는 앵커 데이터와 포지티브 샘플 간의 유사도 산출하는데, 유사도는 두 데이터 간의 특징을 비교하여 얼마나 유사한지를 나타내는 값으로서, 유사도가 높을수록 손실이 감소한다.
- [0102] 분모 부분은 앵커 데이터와 지정된 포지티브 샘플 간의 유사도와 앵커 데이터와 지정된 네거티브 샘플 간의 유사도를 모두 고려하는데, 이는 유사한 클래스 간의 거리를 최소화하고, 다른 클래스 간의 거리를 최대화하여 모델이 데이터 간의 경계를 명확하게 학습할 수 있도록 한다.
- [0103] 근데 이렇게 설계하게 되면, 두 부분의 유사도가 분수로 묶여 있기 때문에 손실이 줄어드는데 두 유사도가 미치는 영향이 서로 다르게 된다.
- [0104] 즉, 분수는 분자에 값이 변화는 것보다 분모의 값이 변화하는 정도에 더 치우쳐져 변화하게 되는 이유에서, 유사도가 미치는 영향이 서로 다르게 된다.
- [0105] 즉, 네거티브 유사도가 줄어들고 포지티브 유사도가 커지는 형태로 손실이 감소하는 것이 아니라 네거티브에 대한 유사도만 줄어드는 형태로 학습될 수 있다.
- [0106] 이에, 학습 처리부(230)에서는 이러한 영향을 줄이기 위해 분수로 묶여 있는 이 두가지 유사도를 분리해서 새로운 대조 손실을 제안하는 것이며, 동일한 멀티모달(intra-modal) 데이터 내에서 연산하는 유사도와, 서로 다른 멀티모달(inter modal) 데이터 간 연산하는 유사도를 각각 독립적으로 산출할 수 있다.
- [0107] 특히, 학습 처리부(230)는 지정된 포지티브 샘플과 앵커 데이터의 유사도는 1이 되도록 하고, 지정된 네거티브 샘플과 상기 앵커 데이터의 유사도는 -1이 되도록 학습할 수 있다.

- [0108] 본 발명은 특정분포가 수렴하지 못해 모델 학습에 방해되는 문제를 방지하며, 분별력 있는 특정분포의 원활한 학습을 가능하게 할 수 있다.
- [0109] 학습 처리부(230)는 어느 하나의 감정에 앵커가 지정되지 않고, 감정에 따라 순차적으로 선정될 수 있도록 한다.
- [0110] 이를 위해, 학습 처리부(230)는 임의의 멀티모달 데이터를 앵커(anchor) 데이터로 지정하되, 멀티모달 데이터의 감정을 기준으로 하여 순차적으로 지정되도록 앵커(anchor) 데이터를 지정할 수 있다.
- [0111] 이를 통해, 본 발명은 샘플 수가 많은 감정에 편향되어 학습되는 것을 방지하고 샘플 수가 적은 감정에 대해서도 준수한 성능을 제공할 수 있고, 다양하고 범용적인 모달리티 데이터 사용과 세밀한 감정 인식을 가능하게 하며, 특히 제한적인 데이터 상황에서 활용 가능성이 높은 감정 인식 기술을 제공할 수 있다.
- [0112] 다음으로, 융합 처리부(240)는 학습된 특징을 융합하고, 감정 인식부(240)는 융합된 데이터를 기반으로 감정을 인식한다.
- [0113] 일실시예에 따르면, 융합 처리부(240)는 모달리티 데이터에 대한 중요도를 기반으로 산출된 가중치를 적용하여 학습된 특징을 융합할 수 있다.
- [0114] 특히 융합 처리부(240)는 완전 연결 층(fully connected layer)으로 구성된 피드 포워드 층(feed forward layer)와 시그모이드(sigmoid) 함수를 적용하여 모달리티 데이터 별 중요도에 따라 0 내지 1사이의 가중치를 학습할 수 있다. 또한, 융합 처리부(240)는 학습된 가중치를 상기 고유 특징(specific feature)과 공통 특징(invariant feature)에 대해 적용할 수 있다.
- [0115] 감정 인식부(250)는 모달리티 데이터 마다 생성되는 고유 특징(specific feature)과 공통 특징(invariant feature)이 감정을 인식하는데 얼마나 기여하는지를 반영해서 감정을 인식할 수 있다.
- [0116] 한편, 제어부(260)는 중앙처리장치(Central Processing Unit, CPU)로 해석될 수 있고, 시스템 내에서 다양한 연산을 수행하고 데이터를 처리하는 역할을 수행할 수 있다.
- [0117] 특히, 제어부(260)는 메모리에서 명령어를 읽어 들이고 해당 명령을 해석하여 실행할 수 있고, 덧셈, 뺄셈, 곱셈, 나눗셈 등의 산술 연산을 수행할 수도 있다.
- [0118] 또한, 제어부(260)는 데이터 저장과 로드 (Data Storage and Retrieval)를 처리할 수 있는데, 메모리에서 데이터를 읽어 들이고, 연산을 수행한 결과를 다시 메모리에 저장하는 기능을 수행할 수도 있다.
- [0119] 또한, 제어부(260)는 프로그램의 실행 흐름을 관리할 수 있는데, 특히, 조건문 (if-else)이나 반복문 (for, while)과 같은 명령을 이용하여 프로그램의 흐름을 제어할 수 있다. 또한, 제어부(260)는 내부에 레지스터라는 작고 빠른 기억장치가 배치될 수 있고, 이 레지스터는 데이터를 임시로 저장하거나 연산을 수행하기 위해 사용될 수도 있다.
- [0120] 제어부(260)는 외부 이벤트나 예외 상황이 발생했을 때 이를 처리하고 적절한 조치를 취할 수 있고, 메인 메모리보다 빠른 캐시 메모리를 사용하여 데이터 및 명령어에 빠르게 액세스할 수 있다.
- [0121] 뿐만 아니라, 제어부(260)는 메모리나 입출력 장치와 통신하기 위해 시스템 버스를 사용할 수 있고, 전력 소모를 최소화하기 위한 다양한 전력 관리 기능을 제공할 수 있다.
- [0122] 도 3은 있도록 트랜스포머 모델의 인코더로 구성된 모달리티 고유 인코더(modality specific encoder)를 통해 각 모달리티마다 모달리티 고유(modality specific) 특징을 생성하는 실시예를 설명하는 도면이다.
- [0123] 모달리티 고유 인코더(modality specific encoder)는 각 modality의 고유한 특징 표현을 학습할 수 있고, 트랜스포머 기반의 서로 다른 인코더를 사용할 수 있다.
- [0124] 또한, 모달리티 공통 인코더(Modality-invariant Encoder)는 모든 모달리티의 공통된 특징표현을 학습할 수 있다.
- [0125] 파라미터를 공유하는 트랜스포머 기반의 동일한 인코더를 사용할 수 있다.
- [0126] 추출된 특징들은 모달리티의 고유한 특징표현을 학습할 수 있도록 트랜스포머 모델의 인코더로 구성된 모달리티 고유 인코더(modality specific encoder)를 통해 각 모달리티마다 모달리티 고유 특징을 생성할 수 있다.
- [0127] 또한 모든 모달리티의 공통된 특징표현을 학습할 수 있도록 파라미터를 공유하는 1개의 트랜스포머 인코더를 통

해 모달리티마다 모달리티의 공통 특징을 생성한다.

- [0128] 본 발명에서는 멀티모달 데이터를 텍스트 데이터, 음성 데이터, 및 생체신호 데이터를 포함하는 것으로 설명한다.
- [0129] 즉, 텍스트 데이터, 음성 데이터, 및 생체신호 데이터의 각각으로부터 추출되는 모달리티 고유 특징들 3개와, 텍스트 데이터, 음성 데이터, 및 생체신호 데이터의 모달리티 공통 특징들 3개의 특징이 생성되어서 총 6개의 특징이 만들어질 수 있다.
- [0130] 도 4는 모달리티의 중요도에 따라 가중하여 이들을 융합하는 융합 처리부에 상응하는 Gated Fusion Block을 설명하는 도면이다.
- [0131] 일실시예에 따른 Gated Fusion Block은 각 트랜스포머 인코더로부터 추출된 특징들을 상호보완적으로 결합할 수 있다.
- [0132] 일실시예에 따른 Gated Fusion Block은 피드포워드 층(Feed Forward layer)과 시그모이드(sigmoid) 함수를 통해 모달리티별로 보다 분별력 있는 특징 강조할 수 있다.
- [0133] 일실시예에 따른 Gated Fusion Block은 앞서 추출한 6개의 특징들에 대해서 모달리티의 중요도에 따라 가중하여 이들을 융합할 수 있다.
- [0134] 일실시예에 따른 Gated Fusion Block은 6개의 특징들에 대해, 첫번째 행에는 고유(specific) 특징을 두번째 행에는 공통(invariant) 특징을 위치시킬 수 있고, 완전 연결층(fully connected layer)으로 구성된 피드포워드 층(Feed Forward layer)과 시그모이드(sigmoid) 함수를 통과하여 가중치를 학습할 수 있다.
- [0135] 이 과정에서, 일실시예에 따른 Gated Fusion Block은 모달리티 별 중요도에 따라 0~1사이의 가중치를 학습할 수 있다. 이 가중치를 기존의 공통(invariant) 특징과 고유(specific) 특징에 곱해주고, 이러한 과정을 모든 모달리티에 수행하게 된다.
- [0136] 일실시예에 따른 Gated Fusion Block은 이 과정을 통해 모달리티마다 생성되는 공통(invariant) 특징과 고유(specific) 특징이 감정을 인식하는데 얼마나 기여하는지를 반영해서 감정을 인식할 수 있다.
- [0137] 도 5a는 각 샘플 내 특징 사이에서 계산하는 위딘 샘플 로스(Within-sample Loss)를 설명하는 도면이다.
- [0138] 위딘 샘플 로스(Within-sample Loss)는 각 샘플 내 특징 사이에서 계산하고, 샘플의 특징 정보를 이용한 학습할 수 있다.
- [0139] 위딘 샘플 로스(Within-sample Loss)는 6가지 특징들의 잠재적인(invariant, specific)과 공통 표현 공간을 학습할 수 있다.
- [0140] 도 5b는 서로 다른 두 샘플들 사이의 특징끼리 계산하는 비트윈 샘플 로스(Between-sample Loss)를 설명하는 도면이다.
- [0141] 비트윈 샘플 로스(Between-sample Loss)는 서로 다른 두 샘플들 사이의 특징끼리 계산할 수 있다.
- [0142] 이 과정에서, 샘플간 레이블 정보를 이용한 학습을 수행하고, Class-discriminative 한 공통 표현 공간을 학습할 수 있다.
- [0143] 본 발명에 따른 멀티모달 데이터 융합 기반의 감정 인식 장치는 위딘 샘플 로스(Within-sample Loss)와 비트윈 샘플 로스(Between-sample Loss)를 이용하여 학습할 수 있다.
- [0144] 즉, 본 발명에 따른 멀티모달 데이터 융합 기반의 감정 인식 장치는 위딘 샘플 로스(Within-sample Loss)를 이용하여 하나의 샘플에서 생성되는 6가지 특징들 사이에서의 연산을 이용하고, 비트윈 샘플 로스(Between-sample Loss)를 이용하여 서로 다른 두 샘플들에서 생성되는 특징들 사이에서 계산된다.
- [0145] 도 6은 위딘 샘플 로스(Within-sample Loss)에서 모달리티-고유(Modality-specific)와 모달리티-공통(Modality-invariant)을 대비하여 설명하는 도면이다.
- [0146] 도면부호 610은 모달리티-고유(Modality-specific)를 학습하되, 텍스트 데이터, 오디오 데이터, 및 생체신호 데이터에 대한 고유 특징(specific feature)에 대해 특징들간 거리가 멀어지도록 학습할 수 있다.
- [0147] 도면부호 620은 모달리티-공통(Modality-invariant)을 학습하되, 텍스트 데이터, 오디오 데이터, 및 생체신호

데이터에 대한 공통 특징(invariant feature)에 대해 거리가 가까워지도록 학습할 수 있다.

- [0148] 본 발명에 따른 멀티모달 데이터 융합 기반의 감정 인식 장치는 위딘 샘플 로스(Within-sample Loss)를 통해 고유 특징(specific feature)들은 서로 다른 모달리티끼리의 거리가 멀어지도록 하여 모달리티의 고유한 정보를 학습하고, 공통 특징(invariant feature)들은 거리가 가까워지도록 하여 모달리티의 공통적인 특징을 학습할 수 있다.
- [0149] 도 7은 위딘 샘플 로스(Within-sample Loss)의 SOC loss와 CMD loss를 설명하는 도면이다.
- [0150] SOC Loss (Soft Orthogonality Constraints Loss)는 모델이 학습하는 동안 특정 제약을 강조하는 손실 함수이다. 여기서 'Soft Orthogonality'는 손실이 모델 파라미터의 각 행렬이 직교(orthogonal) 또는 근사적으로 직교하도록 유도하는 방식을 나타낸다.
- [0151] 직교 행렬은 행과 열 벡터가 서로 수직이며 단위 길이를 갖는 행렬을 의미하는데, 모델에서 이러한 특성을 가진 행렬을 학습하는 것은 유용할 수 있다. 예를 들어, 자연어 처리 분야에서 임베딩 행렬이 서로 수직이면 서로 다른 단어 간의 의미론적 유사성을 잘 캡처할 수 있다.
- [0152] SOC Loss는 이러한 특성을 손실 함수를 통해 강조하고자 하는 목적으로 사용될 수 있다. Soft Orthogonality를 강제하기 위해 일반적으로는 각 행렬의 행 또는 열 간의 내적이 0에 가까워지도록 손실을 설계할 수 있다.
- [0153] 본 발명에서 SOC loss(Soft Orthogonality Constraints Loss)는 동일 모달리티의 공통(invariant) 특징과 고유(specific) 특징 간 거리가 멀어지도록 학습할 수 있고, 서로 다른 모달리티(modality)의 고유(specific) 특징 간 거리가 멀어지도록 학습할 수 있다.
- [0154] Central Moment Discrepancy Loss (CMD Loss)는 두 분포 간의 차이를 측정하고 최소화하기 위한 손실 함수 중 하나로서, 이 손실은 분포 간의 중심적 특징을 고려하여 분포 간의 차이를 평가하는 데 중점을 둘 수 있다.
- [0155] 여기서 중심적 특징은 평균, 분산, 왜도, 첨도와 같은 통계량을 의미하고, 이러한 특징들은 분포의 형태와 모양을 설명하는 중요한 특성에 해당한다.
- [0156] 본 발명에서 Central Moment Discrepancy Loss (CMD)는 모달리티 공통(Modality-invariant) 특징들 간 거리가 가까워지도록 학습할 수 있다.
- [0157] 위딘 샘플 로스(Within-sample Loss) SOC loss와 CMD loss를 포함한다.
- [0158] 즉, SOC loss는 외적 연산을 통해서 두 특징간의 거리가 멀어지도록 하는 식으로써 동일 모달리티의 공통(invariant) 특징과 고유(specific) 특징 간 거리가 멀어지도록 하고 서로 다른 모달리티의 고유(specific) 특징 간 거리 또한 멀어지도록 설계된 함수이다.
- [0159] CMD loss는 두 특징 표현의 분포간 거리를 계산하는 식으로 공통(invariant) 특징들 사이의 거리가 가까워지도록 설계된 함수이다.
- [0160] 이러한 두 함수를 통해 각 모달리티의 고유한 특징을 학습하면서도 공통된 특징을 학습할 수 있다.
- [0161] 도 8은 동일한 감정을 갖는 샘플들의 특징은 유사하도록, 다른 감정을 갖는 샘플들의 특징들은 차이가 나도록 하는 비트윈 샘플 로스(Between-sample Loss)를 설명하는 도면(800)이다.
- [0162] 도면부호 801은 행복한(happy) 감정을 갖는 샘플들이고, 도면부호 802는 슬픈(sad) 감정을 갖는 샘플들을 나타낸다.
- [0163] 비트윈 샘플 로스(Between-sample Loss)는 동일한 감정 레이블을 가진 샘플에 대해서는 유사한 데이터 분포를 형성하도록 하고, 다른 감정 레이블을 가진 샘플에 대해서는 구별되는 데이터 분포를 형성하도록 한다.
- [0164] 본 발명에 따른 비트윈 샘플 로스(Between-sample Loss)는 Class-discriminative 한 공통 표현 공간 학습을 기반으로 유사한 특징을 갖는 샘플끼리 인접하게 하고, 다른 특징을 갖는 샘플끼리는 멀어지게 배치되도록 표현 공간을 학습한다.
- [0165] 비트윈 샘플 로스(Between-sample Loss)는 임의의 샘플을 앵커(anchor) 샘플(또는 앵커)로 지정하고, 지정된 앵커와 같은 클래스를 갖는 샘플을 포지티브 샘플로 지정하고, 앵커와 다른 클래스를 갖는 샘플을 네거티브 샘플로 지정할 수 있다.
- [0166] 이에, 비트윈 샘플 로스(Between-sample Loss)는 앵커와 유사하도록, 네거티브 샘플은 앵커와 차이가 나도록 학

습할 수 있다.

- [0167] 도 9는 배치(batch) 단위로 계산되는 비트윈 샘플 로스(Between-sample Loss)를 설명하는 도면이다.
- [0168] 대조 손실(Contrastive loss)은 배치 단위로 계산될 수 있다.
- [0169] 본 발명에 따른 멀티모달 데이터 융합 기반의 감정 인식 장치는 배치 내에서 하나의 앵커를 선정하고, 포지티브 샘플에 대해서는 당기고(pull), 네거티브 샘플에 대해서는 밀도록(Push)할 수 있다.
- [0170] 멀티모달 데이터 융합 기반의 감정 인식 장치는 대조 손실(Contrastive loss)은 배치 단위로 계산하는데, 도 9의 오른쪽과 같이 앵커로 지정된 샘플의 감정이 평온(neutral)이면 앵커의 감정과 동일하게 평온(neutral) 감정을 나타내는 첫번째 샘플은 포지티브 샘플이 되고, 평온(neutral)과 다른 감정을 나타내는 행복(Happy)와 역겨움(disgust)의 감정을 갖는 샘플은 네거티브 샘플로 지정될 수 있다.
- [0171] 도 10은 본 발명에 따른 contrastive loss로서, 같은 modality특징 간 계산과 다른 modality특징 간 계산을 구분하여 처리하는 Between-sample Loss를 설명하는 도면이다.
- [0172] 학습 처리부는, 동일한 모달리티(intra-modal) 데이터 및 서로 다른 멀티모달(inter modal) 데이터 간 특징들에 대해서는 감정에 따라 거리가 가까워지거나 멀어지도록 학습할 수 있다.
- [0173] 동일한 멀티모달(intra-modal) 데이터는 Intra-modal Contrastive Loss (Intra-CL)로서, 같은 모달리티 특징 간 계산될 수 있고, 서로 다른 멀티모달(inter modal) 데이터는 Inter-modal Contrastive Loss (Inter-CL)로서, 다른 모달리티 특징 간 계산될 수 있다.
- [0174] 본 발명에서는 이러한 대조 손실(Contrastive loss)을 같은 모달리티 내에서 계산하는 Intra-modal contrastive loss와 다른 모달리티 사이에서 계산하는 Inter modal contrastive loss로 나눠서 적용할 수 있다.
- [0175] 기존의 일반적인 대조 손실(Contrastive loss)은 분자에는 앵커와 포지티브 샘플 간의 유사도를 구하고, 분모에는 앵커와 포지티브 샘플 간 유사도와 앵커와 네거티브 샘플 간의 유사도를 구할 수 있다.
- [0176] 이렇게 손실(loss)을 설계하는 경우 앵커와 포지티브 샘플 간의 유사도가 크면 클수록, 앵커와 네거티브 샘플 간의 유사도가 작으면 작을수록 로스가 줄어들게 된다.
- [0177] 근데 이렇게 설계하게 되면, 이 두 유사도가 분수로 묶여 있기 때문에 로스가 줄어드는데 두 유사도가 미치는 영향이 서로 다르게 된다.
- [0178] 일반적으로 분수는 분자에 값이 변화는 것보다 분모의 값이 변화하는 정도에 더 치우쳐져 변화하게 되는데, 네거티브 유사도가 줄어들고 포지티브 유사도가 커지는 형태로 로스가 감소하는 것이 아니라 네거티브에 대한 유사도만 줄어드는 형태로 학습될 수 있다.
- [0179] 본 발명에서는, 이러한 영향을 줄이기 위해 분수로 묶여 있는 이 두가지 유사도를 분리해서 새로운 대조 손실(Contrastive loss)을 설계하였다.
- [0180] 일실시예에 따른 멀티모달 데이터 융합 기반의 감정 인식 장치에서는 추가적으로 단순히 샘플들 사이에서만 이 손실을 계산하는 것이 아니라 앞서 생성한 공통 특징(invariant feature)과 고유 특징(specific feature)들의 모달리티 별 유사도도 고려할 것이기 때문에 손실의 수정이 필요하다.
- [0182] 본 발명에 따른 멀티모달 데이터 융합 기반의 감정 인식 장치는 아래 [수학식 1]을 이용하여 Intra-modal Contrastive Loss (Intra-CL)를 산출한다.

[0184] [수학식 1]

$$E_m \left(\text{sim}(a_m, p_m) \right) = \frac{1}{6} \left[\text{sim}(a_{V_t}, p_{V_t}) + \text{sim}(a_{V_a}, p_{V_a}) + \text{sim}(a_{V_b}, p_{V_b}) + \text{sim}(a_{S_t}, p_{S_t}) + \text{sim}(a_{S_a}, p_{S_a}) + \text{sim}(a_{S_b}, p_{S_b}) \right]$$

$$L_{\text{Intra-CL}} = \left[1 - E_P \left\{ E_m \left(\text{sim}(a_m, p_m^i) \right) \right\} \right] + \left[1 + E_N \left\{ E_m \left(\text{sim}(a_m, n_m^j) \right) \right\} \right]$$

$$m \in V_t, V_a, V_b, S_t, S_a, S_b$$

[0185]

[0187] [수학식 1]에서 은 모달리티, sim은 유사도를 나타낸다.

[0188] P 는 배치 내의 포지티브 샘플의 개수이며 N 은 네거티브 샘플의 개수이다.

[0189] a_m 는 배치 내 선정된 anchor 이고, p_m^i 는 positive pair 들 중 번째, n_m^j 는 negative pair 들 중 번째 특징(features)이다.

[0190] Intra-CL 은 동일 모달리티 내에서 계산되며, Inter-CL은 modality-invariant 특징에 한정하여 다른 모달리티 사이에서 계산된다.

[0191] V는 세 가지 모달리티가 공유하는 특징 공간 상에 위치한 벡터를 의미한다. V_t 는 텍스트 데이터에 대한 특징벡터, V_a 는 오디오 데이터에 대한 특징벡터, V_b 는 생체신호 데이터에 대한 특징벡터를 나타낸다.

[0192] S는 세 가지 모달리티가 독립적으로 가지는 특징 공간 상에 위치한 벡터를 의미한다. S_t 는 텍스트 데이터에 대한 특징벡터, S_a 는 오디오 데이터에 대한 특징벡터, S_b 는 생체신호 데이터에 대한 특징벡터를 나타낸다.

[0194] 또한, 본 발명에 따른 멀티모달 데이터 융합 기반의 감정 인식 장치는 아래 [수학식 2]을 이용하여 오직 공통 특징(Invariant feature)에 대해서만, Inter-modal Contrastive Loss (Inter-CL)를 산출한다.

[0196] [수학식 2]

$$E_m \left(\text{sim}(a_{m_1}, p_{m_2}) \right) = \frac{1}{6} \left[\text{sim}(a_{V_t}, p_{V_a}) + \text{sim}(a_{V_t}, p_{V_b}) + \text{sim}(a_{V_a}, p_{V_t}) + \text{sim}(a_{V_a}, p_{V_b}) + \text{sim}(a_{V_b}, p_{V_t}) + \text{sim}(a_{V_b}, p_{V_a}) \right]$$

$$L_{\text{Inter-CL}} = \left[1 - E_P \left\{ E_m \left(\text{sim}(a_{m_1}, p_{m_2}^i) \right) \right\} \right] + \left[1 + E_N \left\{ E_m \left(\text{sim}(a_{m_1}, n_{m_2}^j) \right) \right\} \right]$$

$$(m_1, m_2) \in \{(V_t, V_a), (V_t, V_b), (V_a, V_b)\}$$

[0197]

[0199] [수학식 2]에서 은 모달리티, sim은 유사도를 나타낸다.

[0200] P 는 배치 내의 포지티브 샘플의 개수이며 N 은 네거티브 샘플의 개수이다.

[0201] a_m 는 배치 내 선정된 anchor 이고, p_m^i 는 positive pair 들 중 번째, n_m^j 는 negative pair 들 중 번째 특징(features)이다.

[0202] Intra-CL 은 모달리티 내에서 계산되며, Inter-CL은 modality-invariant 특징에 한정하여 다른 모달리티 사이에서 계산된다.

[0203] V는 세 가지 모달리티가 공유하는 특징 공간 상에 위치한 벡터를 의미한다. V_t 는 텍스트 데이터에 대한 특징벡터, V_a 는 오디오 데이터에 대한 특징벡터, V_b 는 생체신호 데이터에 대한 특징벡터를 나타낸다.

[0204] S는 세 가지 모달리티가 독립적으로 가지는 특징 공간 상에 위치한 벡터를 의미한다. S_t 는 텍스트 데이터에 대한 특징벡터, S_a 는 오디오 데이터에 대한 특징벡터, S_b 는 생체신호 데이터에 대한 특징벡터를 나타낸다.

- [0206] 먼저, [수학식 1]로 표현되는 Intra-modal contrastive loss는 동일 모달리티 내에서 연산하는 손실로서, 모달리티 마다의 공통(invariant) 특징과 고유(specific) 특징을 포함하는 총 6가지 특징들에 대해 앵커와의 유사도를 통해 앵커와 동일한 감정을 갖는 샘플이라면 유사해지도록 학습시킬 수 있다.
- [0207] 즉, 포지티브 샘플과 앵커의 유사도는 1이 되도록 하고, 네거티브 샘플과 앵커의 유사도는 -1이 되도록 할 수 있다.
- [0208] 특히, 앵커의 텍스트 데이터에 대한 공통 특징과 포지티브 샘플의 텍스트 데이터에 대한 공통 특징의 유사도를 계산하고, 오디오 데이터와 생체신호 데이터도 동일하게 계산할 수 있다. 또한, 고유(Specific) 특징에 대해서도 마찬가지로 유사도를 계산할 수 있다.
- [0209] 다음으로 본 발명에 따른 멀티모달 데이터 융합 기반의 감정 인식 장치는 Inter modal contrastive loss를 이용하여 서로 다른 모달리티 사이에서 연산할 수 있다.
- [0210] 이 손실은 공통(Invariant) 특징에만 적용될 수 있다.
- [0211] 예를 들어, 고유(Specific) 특징은 모달리티의 고유한 정보가 담겨 있기 때문에 서로 다른 모달리티 사이에서 연산하는 이 Inter modal contrastive loss를 적용할 경우 이에 대한 정보가 훼손될 수 있어서 공통(Invariant) 특징에만 적용할 수 있다.
- [0212] 포지티브 샘플과 앵커의 서로 다른 모달리티 사이에서의 유사도는 1이 되도록하고, 네거티브 샘플과 앵커의 서로 다른 모달리티 사이에서의 유사도는 -1이 되도록 하는 식입니다.
- [0213] [수학식 2]를 살펴보면, 앵커의 텍스트에 대한 공통(Invariant) 특징과 포지티브 샘플의 오디오에 대한 공통(Invariant) 특징의 유사도를 구할 수 있다.
- [0214] 또한, 앵커의 텍스트 데이터에 대한 공통(Invariant) 특징과 포지티브 샘플의 생체신호 데이터에 대한 공통(Invariant) 특징의 유사도를 구하고, 마찬가지로 앵커의 오디오 데이터에 대한 공통(Invariant) 특징과 생체신호 데이터의 공통(Invariant) 특징들의 유사도를 계산할 수 있다.
- [0215] 도 11은 감정에 따라 순차적으로 앵커가 선정되도록 하는 비트윈 샘플 로스(Between-sample Loss)를 설명하는 도면(1100)이다.
- [0216] 본 발명에 따른 멀티모달 데이터 융합 기반의 감정 인식 장치는 학습 처리부에서 임의의 멀티모달 데이터를 앵커(anchor) 데이터로 지정하되, 멀티모달 데이터의 감정을 기준으로 하여 순차적으로 지정되도록 앵커(anchor) 데이터를 지정할 수 있다.
- [0217] 본 발명에 따른 멀티모달 데이터 융합 기반의 감정 인식 장치는 데이터 불균형으로 특정 감정 레이블만 반복적으로 앵커에 선정될 수 있다.
- [0218] 구체적으로, 배치내에 임의의 샘플을 랜덤하게 앵커로 지정하는 과정에서, 특정 감정에 대한 샘플이 많다면 대부분의 배치에서 동일한 감정을 갖는 샘플이 앵커로 선정될 수 있다.
- [0219] 이를 해소하기 위해, 본 발명에서는 멀티모달 데이터의 감정을 기준으로 하여 순차적으로 지정되도록 앵커(anchor) 데이터를 지정할 수 있다.
- [0220] 이를 통해, 데이터가 불균형한 상황에서도 모든 감정에 대한 대조 손실이 골고루 적용될 수 있다.
- [0221] 도 12는 감정 인식을 위한 멀티모달 융합 학습 네트워크를 설명하는 도면(1200)이다.
- [0222] 설계한 대조 손실(1280)은 모달리티 인코더(1240) 이후에 위치하여, Gated fusion block(1270)을 통과한 후 최종적으로 감정을 인식할 수 있다.
- [0223] 구체적으로, 멀티모달 데이터(1210)는 싱글모달 데이터 보다 더 다양한 정보를 제공할 수 있다. 그러나, 모달리티별 고유한 특성과 표현방식으로 멀티모달 데이터(1210) 간 이질성 문제 해결 필요하다.
- [0224] 본 발명에 따른 멀티모달 데이터 융합 기반의 감정 인식 장치는 텍스트 데이터, 음성 데이터, 또는 생체신호 데이터의 멀티모달 데이터(1210)를 수집하고 이를 가공하여 융합할 수 있다.
- [0225] 멀티모달 데이터 융합 기반의 감정 인식 장치는 멀티모달 데이터(1210)를 융합하는데, 각 모달리티 데이터의 감

재적인 정보를 활용하기 위해, 도면부호 (1220)과 같이 각 모달리티 데이터로부터 사전학습된 모델이나 임의로 계산한 특징들(1230)을 추출한다.

- [0226] 또한, 트랜스포머 인코더(1240)를 통해 모달리티 별 특화된 특징들과 모달리티의 공통적으로 드러나는 특징들을 학습할 수 있다.
- [0227] 트랜스포머 인코더(1240)의 출력은 모달리티 데이터 별로 고유 특징(specific feature) 및 상기 모달리티 데이터의 공통 특징(invariant feature)에 해당한다.
- [0228] 이후 이러한 특징들을 잘 융합시키기 위해 Gated fusion block(1270)을 통과 하여 최종적으로 감정을 인식하게 된다.
- [0229] 설계한 대조 손실(1280)은 모달리티 인코더(1240) 이후에 위치하여, Gated fusion block(1270)을 통과한 후 최종적으로 감정을 인식할 수 있다.
- [0230] 결국, 모달리티별 상호 보완적인 특징을 추출하고 이들을 융합시키기 위한 네트워크를 설계하고, 기존의 contrastive loss를 개선하여 불균형한 데이터 환경에서도 모든 감정에 대해 고루 예측할 수 있는 감정 인식 모델을 제시할 수 있다.
- [0231] 또한, 본 발명에서는 모달리티 데이터 특성에 맞게 추출된 특징들은 공통된 특징 공간에 위치시키는 modality-invariant 특징과 모달리티별로 고유한 특성을 나타내는 modality-specific 특징을 생성하는 transformer 기반의 Modality Representations Encoder를 통과한다. 추출된 특징들은 Gated Fusion Block을 거쳐 융합되며 이를 기반으로 감정을 예측할 수 있다.
- [0232] 도 13은 일실시예에 따른 멀티모달 데이터 융합 기반의 감정 인식 방법을 설명하는 도면이다.
- [0233] 일실시예에 따른 멀티모달 데이터 융합 기반의 감정 인식 방법은 복수의 모달리티 데이터를 포함하는 멀티모달 데이터를 수집할 수 있다(단계 1301).
- [0234] 일실시예에 따른 멀티모달 데이터 융합 기반의 감정 인식 방법은 복수의 모달리티 데이터를 포함하는 멀티모달 데이터를 수집하는 과정에서, 텍스트 데이터, 음성 데이터, 또는 생체신호 데이터 중에서 적어도 둘 이상의 모달리티 데이터를 수집할 수 있다. 특히, 생체신호 데이터는 피부전도도, 손목 피부온도, 맥파 데이터에 대한 시간, 주파수, 시간-주파수 도메인별 특징을 포함할 수 있다.
- [0235] 또한, 일실시예에 따른 멀티모달 데이터 융합 기반의 감정 인식 방법은 사전에 학습된 모델들을 이용하여, 상기 복수의 모달리티 데이터 각각으로부터 복수의 특징정보를 추출할 수 있다.
- [0236] 이 과정에서, 각각의 모달리티 데이터 별로 고유 특징(specific feature) 및 모달리티 데이터의 공통 특징(invariant feature)을 구분하여 추출할 수 있다(단계 1302).
- [0237] 일실시예에 따른 멀티모달 데이터 융합 기반의 감정 인식 방법은 추출된 복수의 특징정보를 학습하되, 각각의 모달리티 데이터 별로 고유한 특징 및 모달리티 데이터의 공통된 특징을 이용하여 학습할 수 있다(단계 1303).
- [0238] 또한, 일실시예에 따른 멀티모달 데이터 융합 기반의 감정 인식 방법은 학습된 특징을 융합하고, 상기 융합된 데이터를 기반으로 감정을 인식할 수 있다(단계 1304).
- [0239] 본 발명에서는 학습된 특징을 융합하고, 상기 융합된 데이터를 기반으로 감정을 인식하기 위해, 모달리티 데이터에 대한 중요도를 기반으로 산출된 가중치를 적용하여 상기 학습된 특징을 융합할 수 있다.
- [0240] 또한, 학습된 특징을 융합하고, 상기 융합된 데이터를 기반으로 감정을 인식하기 위해, 완전 연결 층(fully connected layer)으로 구성된 피드 포워드 층(feed forward layer)와 시그모이드(sigmoid) 함수를 적용하여 모달리티 데이터 별 중요도에 따라 0 내지 1사이의 가중치를 학습할 수 있다. 또한, 학습된 가중치를 상기 고유 특징(specific feature)과 상기 공통 특징(invariant feature)에 대해 적용하고, 모달리티 데이터 마다 생성되는 고유 특징(specific feature)과 상기 공통 특징(invariant feature)이 감정을 인식하는데 얼마나 기여하는지를 반영해서 감정을 인식할 수 있다.
- [0241] 본 발명에 따르면, 추출된 복수의 특징정보를 학습하기 위해, 서로 다른 모달리티 데이터의 고유 특징(specific feature)에 대해서는 거리가 멀어지도록 학습하고, 공통 특징(invariant feature)에 대해서는 거리가 가까워지도록 학습할 수 있다. 동일한 모달리티 데이터의 고유 특징(specific feature)와 공통 특징(invariant feature) 사이에 대해서는 거리가 멀어지도록 학습할 수 있다.

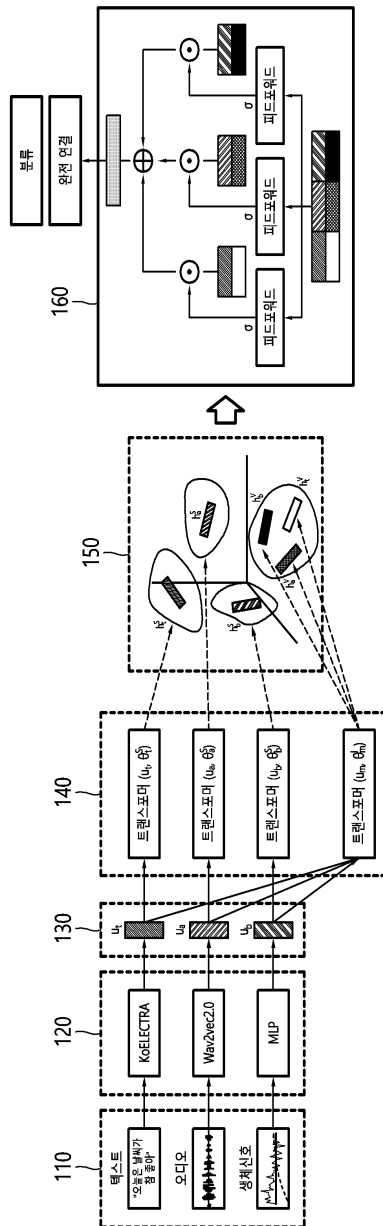
- [0242] 본 발명에 따르면, 추출된 복수의 특징정보를 학습하기 위해, 임의의 멀티모달 데이터를 앵커(anchor) 데이터로 지정하고, 앵커 데이터와 같은 클래스를 갖는 멀티모달 데이터에 대해서는 포지티브 샘플로 지정하고, 상기 앵커 데이터와 다른 클래스를 갖는 멀티모달 데이터에 대해서는 네거티브 샘플로 지정할 수 있다.
- [0243] 또한, 본 발명에서는 대비 손실(contrastive loss)을 활용하여 지정된 포지티브 샘플에 대해서는 앵커 데이터와 유사하도록 학습하고, 상기 지정된 네거티브 샘플에 대해서는 앵커 데이터와 구별되도록 학습할 수 있다.
- [0244] 본 발명에서는 지정된 포지티브 샘플과 상기 앵커 데이터의 유사도는 1이 되도록 하고, 지정된 네거티브 샘플과 상기 앵커 데이터의 유사도는 -1이 되도록 학습할 수 있다. 또한, 대비 손실(contrastive loss)을 활용하는 과정에서, 동일한 멀티모달(intra-modal) 데이터 내에서 연산하는 유사도와, 서로 다른 멀티모달(inter modal) 데이터 간 연산하는 유사도를 각각 독립적으로 산출하되, 상기 동일한 멀티모달(intra-modal) 데이터 내에서는 고유 특징(specific feature)과 공통 특징(invariant feature)을 모두 적용하여 유사도를 연산하고, 서로 다른 멀티모달(inter modal) 데이터 간에는 공통 특징(invariant feature)만을 적용하여 유사도를 연산할 수 있다.
- [0245] 일실시예에 따르면, 임의의 멀티모달 데이터를 앵커(anchor) 데이터로 지정하기 위해, 임의의 멀티모달 데이터를 앵커(anchor) 데이터로 지정할 수 있다. 또한, 멀티모달 데이터의 감정을 기준으로 하여 순차적으로 지정되도록 상기 앵커(anchor) 데이터를 지정할 수 있다.
- [0246] 결국, 본 발명을 이용하면, 모달리티별 상호 보완적인 특징을 추출하고 이들을 융합시키기 위한 네트워크를 설계하고, 기존의 contrastive loss를 개선하여 불균형한 데이터 환경에서도 모든 감정에 대해 고루 예측할 수 있는 감정 인식 모델을 제시할 수 있다.
- [0247] 또한, 본 발명을 이용하면 모달리티 데이터 특성에 맞게 추출된 특징들은 공통된 특징 공간에 위치시키는 modality-invariant 특징과 모달리티별로 고유한 특성을 나타내는 modality-specific 특징을 생성하는 transformer 기반의 Modality Representations Encoder를 통과한다. 추출된 특징들은 Gated Fusion Block을 거쳐 융합되며 이를 기반으로 감정을 예측할 수 있다.
- [0248] 뿐만 아니라, 특징분포가 수렴하지 못해 모델 학습에 방해되는 문제를 방지하며, 분별력 있는 특징분포의 원활한 학습을 가능하게 할 수 있고, 샘플 수가 많은 감정에 편향되어 학습되는 것을 방지하고 샘플 수가 적은 감정에 대해서도 준수한 성능을 제공할 수 있다.
- [0249] 또한, 본 발명을 이용하면 다양하고 범용적인 모달리티 데이터 사용과 세밀한 감정 인식을 가능하게 하며, 특히 제한적인 데이터 상황에서 활용 가능성이 높은 감정 인식 기술을 제공할 수 있다.
- [0250] 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 콘트롤러를 포함할 수 있다. 또한, 병렬 프로세서(parallel processor)와 같은, 다른 처리 구성(processing configuration)도 가능하다.
- [0251] 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 장치, 또는 전송되는 신호 파(signal wave)에 영구적으로, 또는 일시적으로 구체화(embody)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.

- [0252] 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. 상기된 하드웨어 장치는 실시예의 동작을 수행하기 위해 하나 이상의 소프트웨어 모듈로서 작동하도록 구성될 수 있으며, 그 역도 마찬가지이다.
- [0253] 이상과 같이 실시예들이 비록 한정된 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.
- [0254] 그러므로, 다른 구현들, 다른 실시예들 및 특허청구범위와 균등한 것들도 후술하는 특허청구범위의 범위에 속한다.

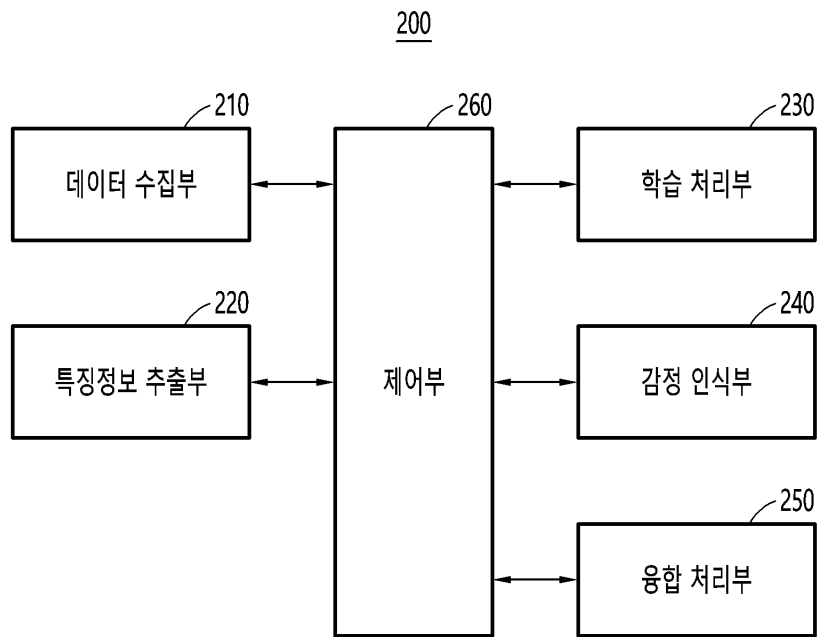
도면

도면1

100

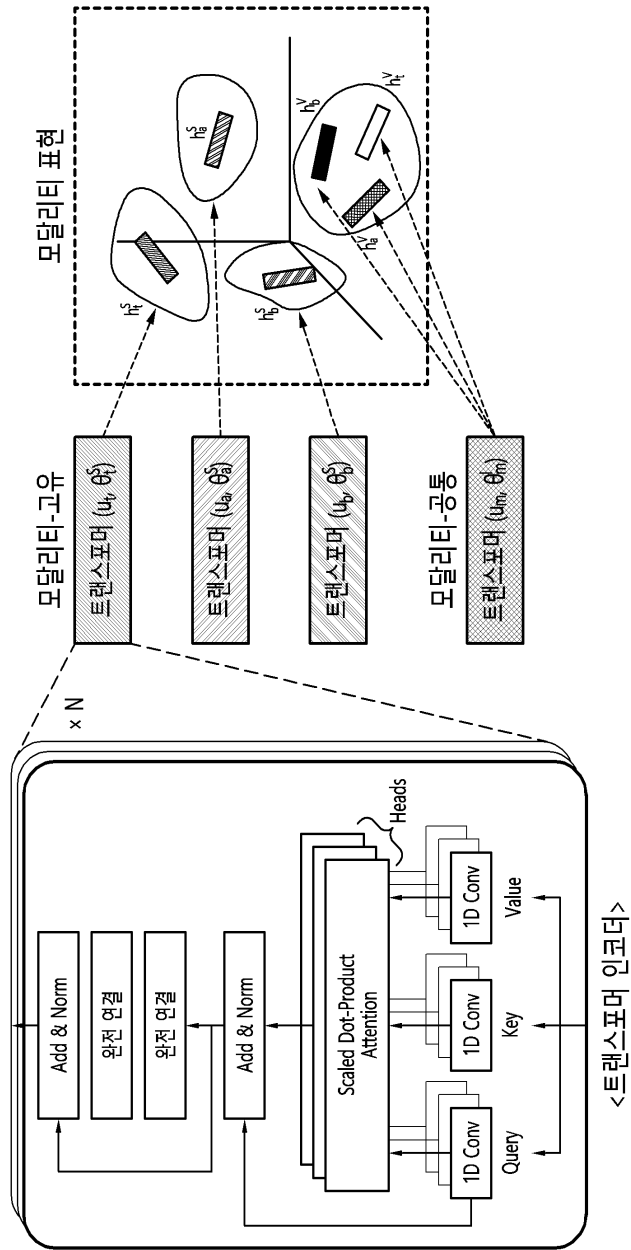


도면2

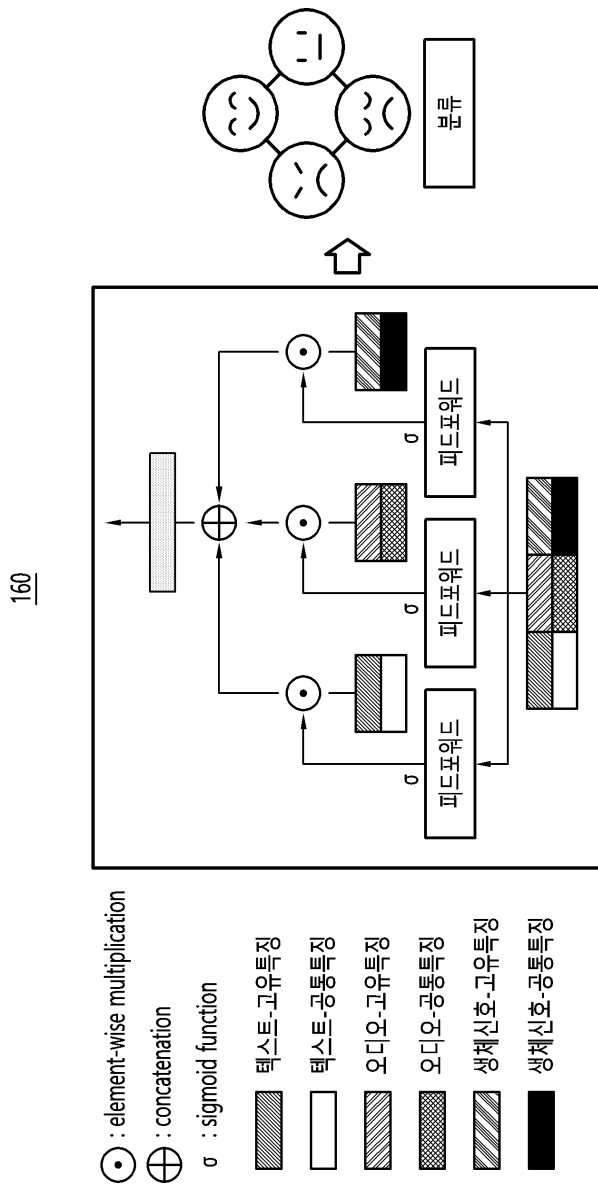


도면3

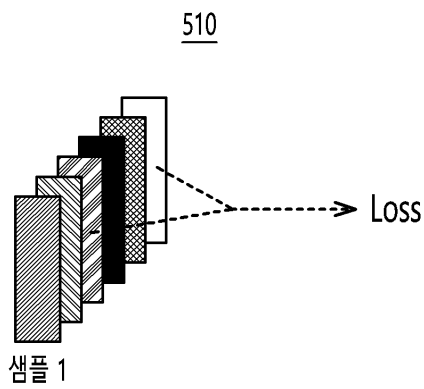
300



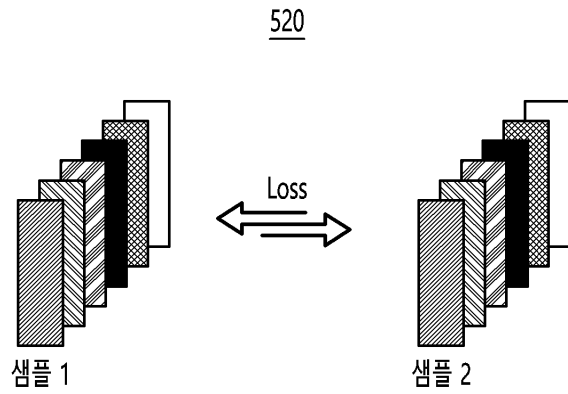
도면4



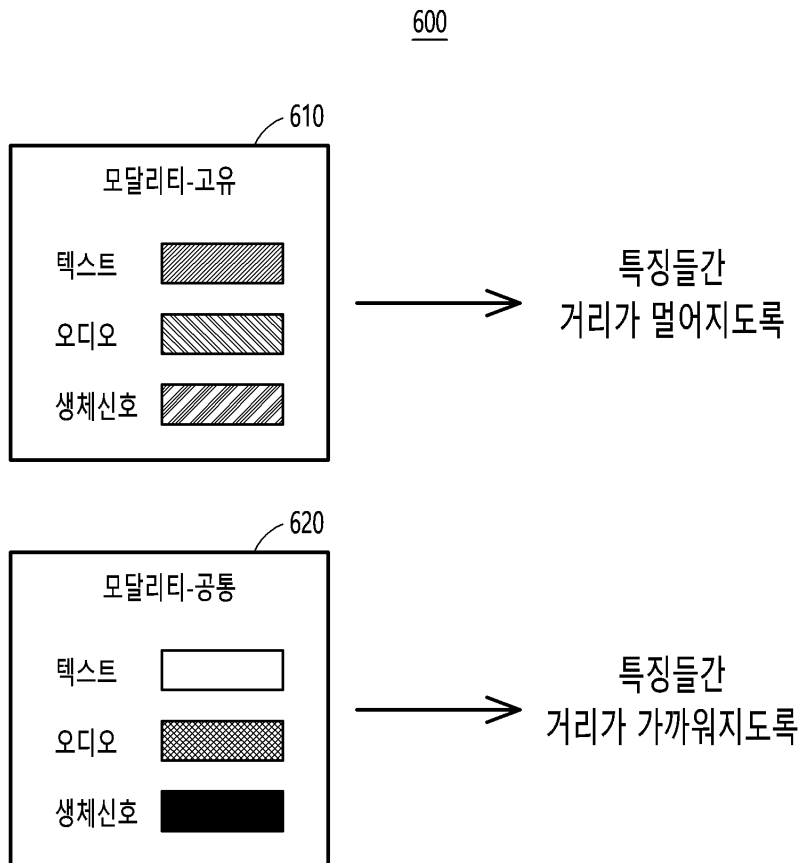
도면5a



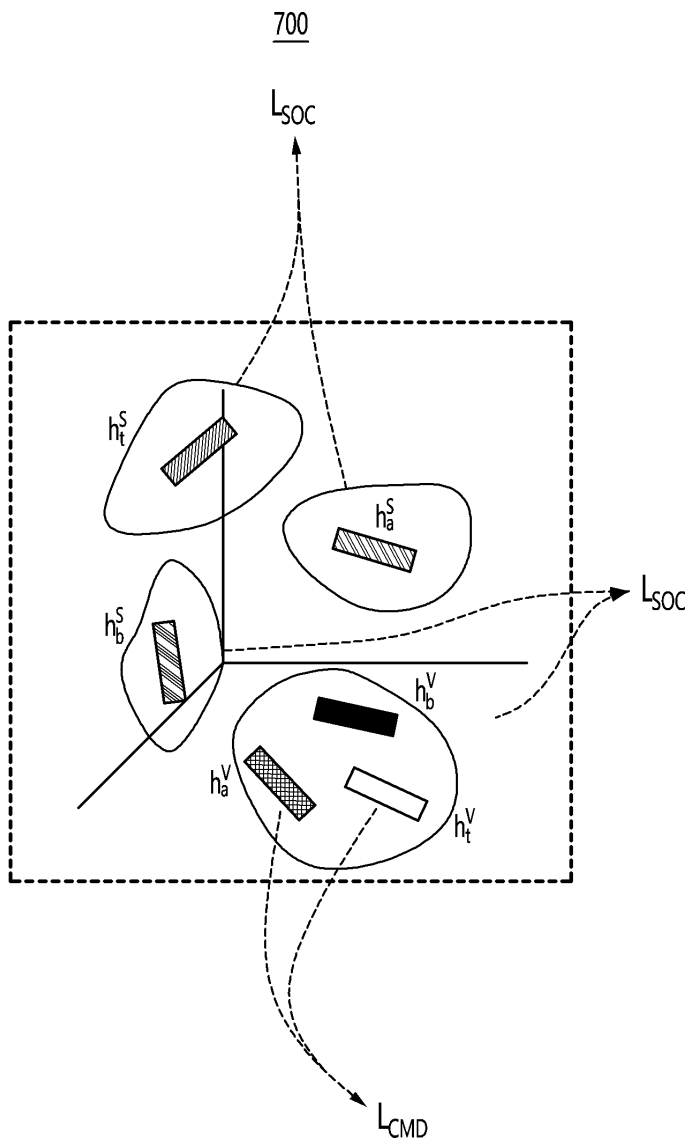
도면5b



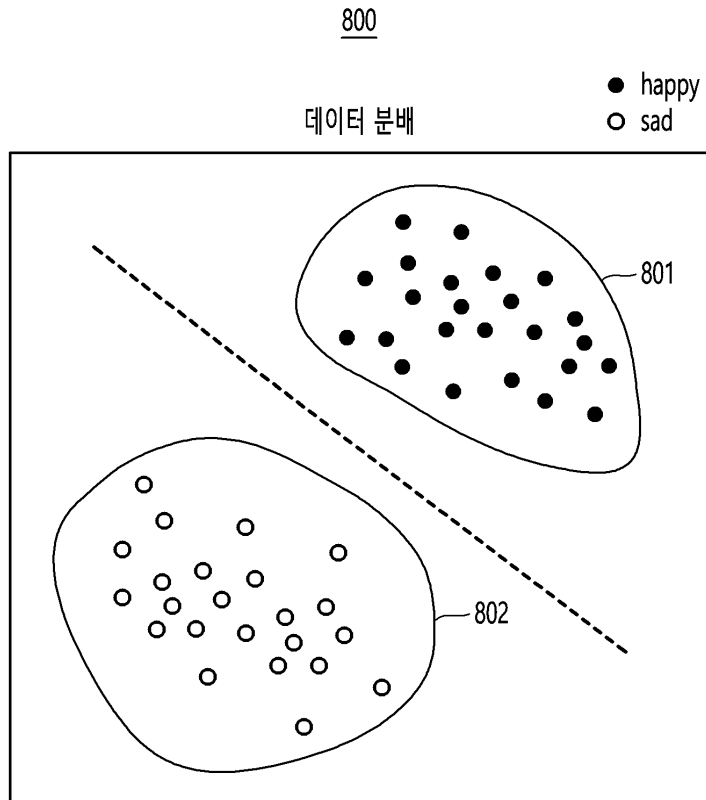
도면6



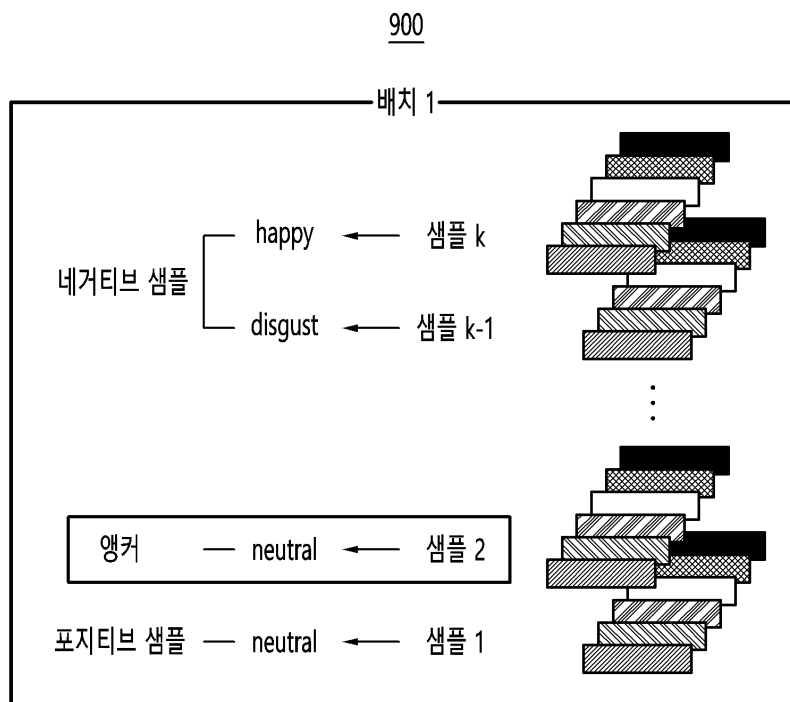
도면7



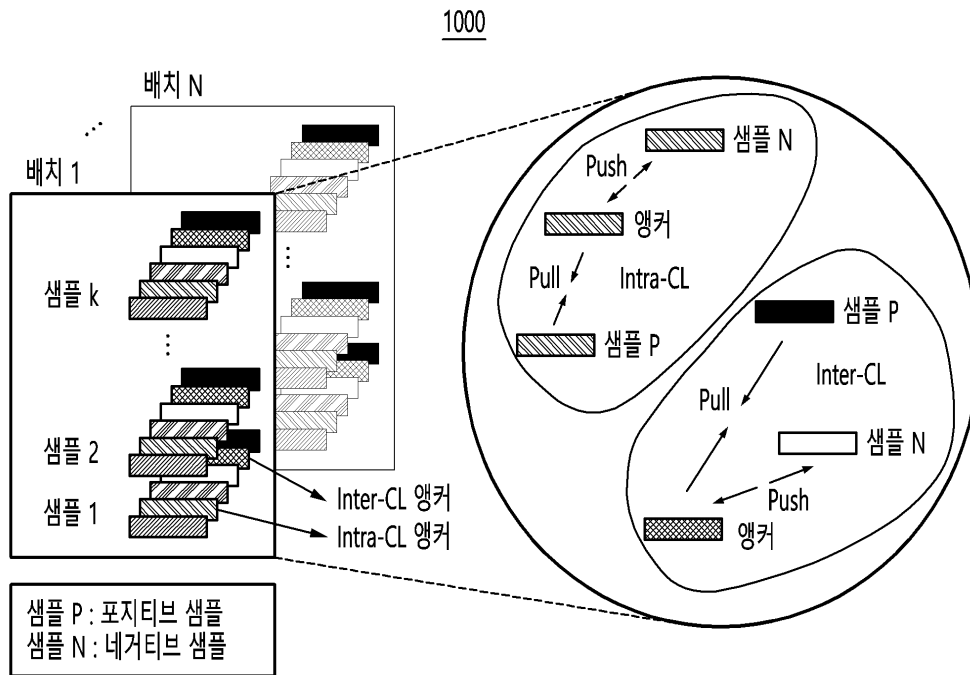
도면8



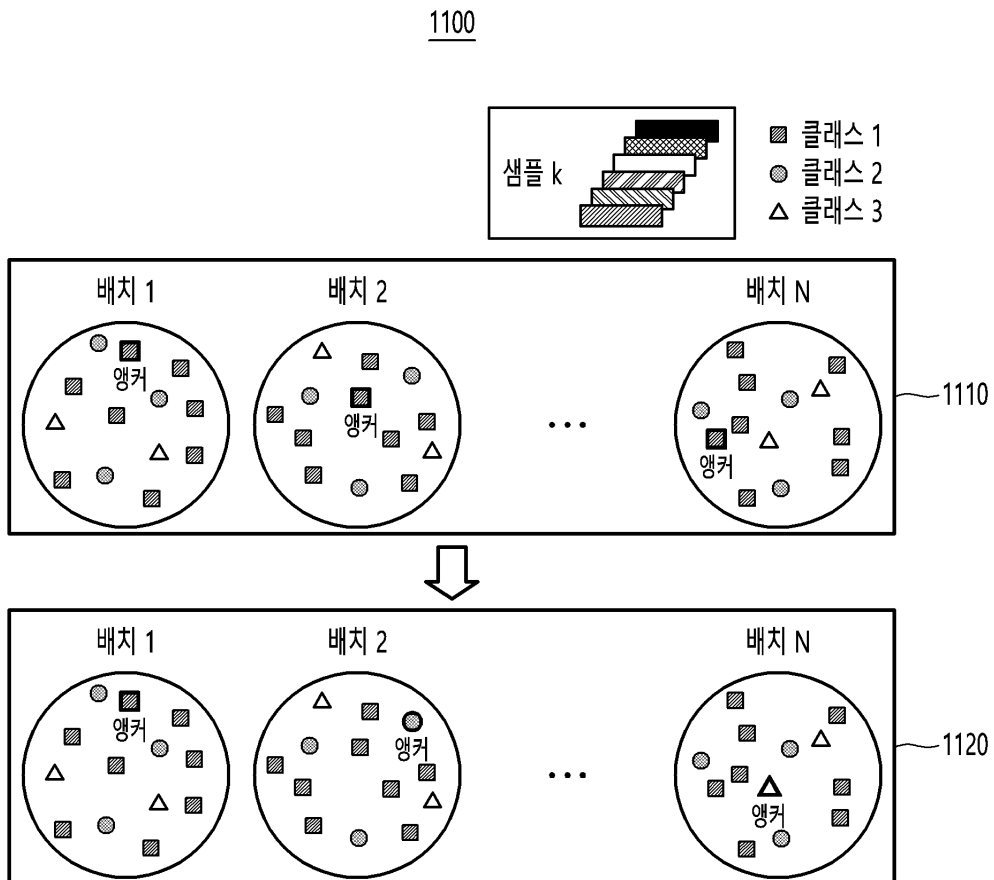
도면9



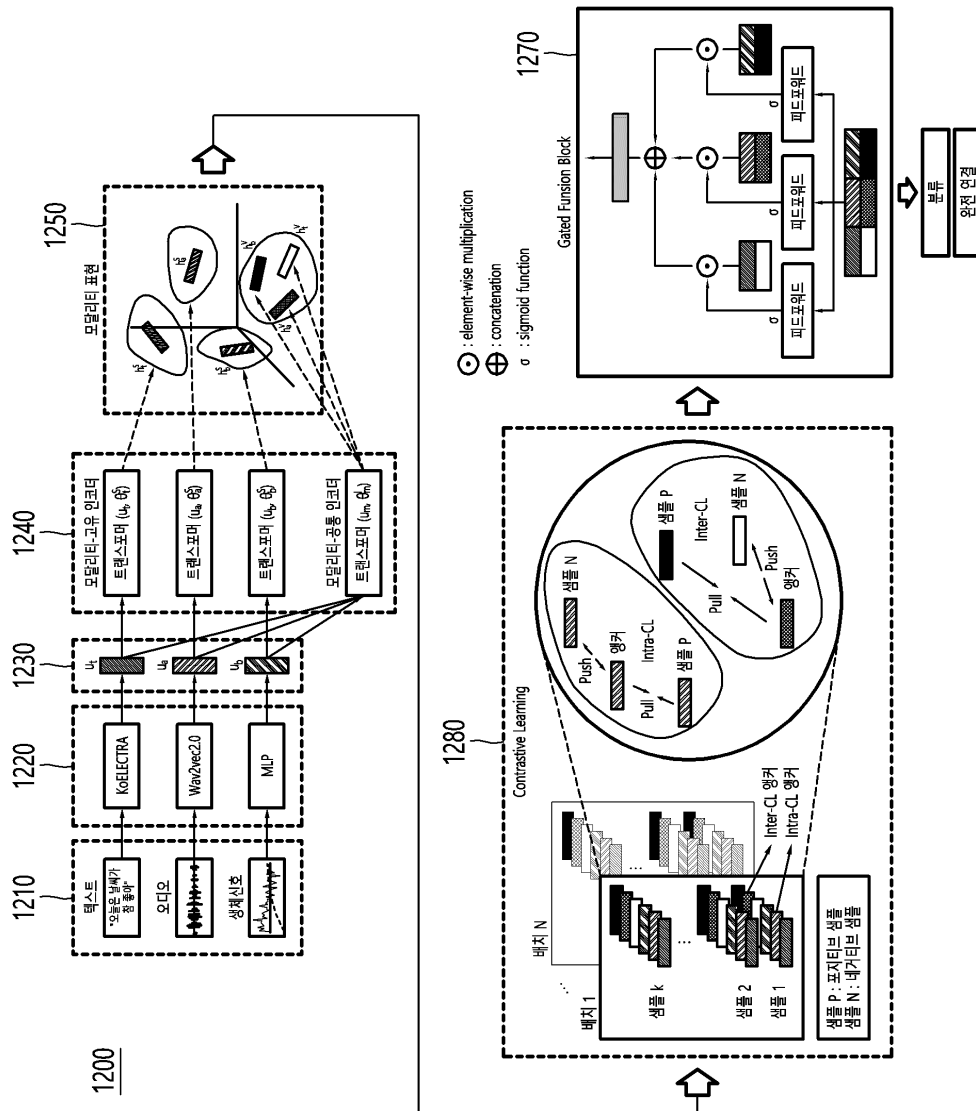
도면10



도면11



도면 12



도면13

