감성 컴퓨팅 논문 최근 연구 동향

A Systematic Review on Affective Computing: Emotion Models, Databases, and Recent Advances

Yan Wang [a], Wei Song [c], Wei Tao [a], Antonio Liotta [d], Dawei Yang [a], Xinlei Li [a], Shuyong Gao [b], Yixuan Sun [a], Weifeng Ge [b], Wei Zhang [b], and Wenqiang Zhang [a,b,*]

[a] Academy for Engineering & Technology, Fudan University, Shanghai 200433; Shanghai Engineering Research Center of AI & Robotics, Shanghai 200433, China; and Engineering Research Center of AI & Robotics, Ministry of Education, Shanghai 200433, China;
yanwang19@fudan.edu.cn (Y.W.); 18110860008@fudan.edu.cn (W.T.); 18110860061@fudan.edu.cn (W.D.); 18110860019@fudan.edu.cn (X.L.); 1609271386@qq.com (Y.S.); wqzhang@fudan.edu.cn (W.Q.);

[b] Shanghai Key Laboratory of Intelligent Information Processing, Fudan University, Shanghai, China;
18110240022@fudan.edu.cn (S.G.); wfge@fudan.edu.cn (W.G.); weizh@fudan.edu.cn (W.Z.); wqzhang@fudan.edu.cn (W.Q.)

[c] College of Information Technology, Shanghai Ocean University, Shanghai 201306, China;
wsong@shou.edu.cn (W.S.)

[d] Faculty of Computer Science, Free University of Bozen-Bolzano, Italy;
antonio.liotta@unibz.it (A.L.)

* Correspondence: wqzhang@fudan.edu.cn; Tel: +86-185-0213-9010

Abstract: With the rapid development of artificial intelligence and the universal promotion of real-life computer applications, affective computing plays a key role in human-computer interactions, entertainment, teaching, safe driving, and multimedia integration. Major breakthroughs have been made recently in the areas of affective computing (i.e., emotion recognition and sentiment analysis). Affective computing is realized based on unimodal or multimodal data, primarily consisting of physical information (e.g., textual, audio, and visual data) and physiological signals (e.g., EEG and ECG signals). Physical-based affect recognition caters to more researchers due to multiple public databases. However, it is hard to reveal one's inner emotion hidden purposely from facial expressions, audio tones, body gestures, etc. Physiological signals can generate more precise and reliable emotional results; yet, the difficulty in acquiring physiological signals also hinders their practical application. Thus, the fusion of physical information and physiological signals can provide useful features of emotional states and lead to higher accuracy. Instead of focusing on one specific field of affective analysis, we systematically review recent

1. 연구 배경

감성 컴퓨팅은 인간의 감정, 정서, 느낌을 컴퓨터가 인식하고 반응하는 능력을 다루는 분야입니다. 1997년 피카드 교수가 이 개념을 제안한 이후, 인공지능의 급속한 발전과 함께 인간-컴퓨터 상호작용, 오락, 교육, 안전 운전 등 다양한 실생활 응용 분야에서 핵심적인 역할을 하고 있습니다.

특히, 사람의 감정은 얼굴 표정(55%), 목소리(38%), 언어(7%)로 표현되지만, 때로는 감정을 의도적으로 숨기는 '사회적 가장'이라는 문제가 있습니다. 이와 달리 EEG나 ECG와 같은 생리적 신호는 감정을 숨기기 어렵기 때문에 더 정확하고 신뢰성 높은 결과를 얻을 수 있습니다. 따라서 물리적 데이터와 생리적 신호를 결합한 **멀티모달(multimodal) 접근법**이 더 높은 정확도를 보여주며 주목받고 있습니다.

2. 연구 목적

이 논문의 목적은 감성 컴퓨팅 분야의 최근 발전을 체계적으로 검토하고, 이 분야의 학계 및 산업 연구자들이 최신 동향을 이해하는 데 도움을 주기 위함입니다. 연구는 단순히 한 분야에 초점을 맞추는 대신, 충분한 데이터와 잘 설계된 방법론이라는 두 가지 측면을 모두 다루고자 합니다. 이를 통해 단일 모달 및 멀티모달 감성 분석의 방법론을 분류하고, 관련 내용을 포괄적으로 제시하는 것을 목표로 합니다.

3. 연구 내용

이 논문은 다음과 같은 구성으로 감성 컴퓨팅을 체계적으로 검토합니다.

- **감정 모델**: 이산적 감정 모델과 다차원적 감정 모델을 소개합니다.
- **데이터베이스**: 감성 컴퓨팅 알고리즘 훈련 및 테스트에 사용되는 네 가지 주요 데이터베이스를 자세히 다룹니다.
- **단일 모달 감정 인식**: 텍스트, 음성, 시각 및 생리적 신호 기반의 감정 인식 방법을 포괄적으로 검토합니다.
- **멀티모달 감성 분석**: 다양한 모달리티를 융합하는 최신 방법론을 심층적으로 분석합니다.
- **논의 및 결론**: 현재의 어려움과 문제점, 그리고 유망한 미래 연구 방향을 제시합니다.

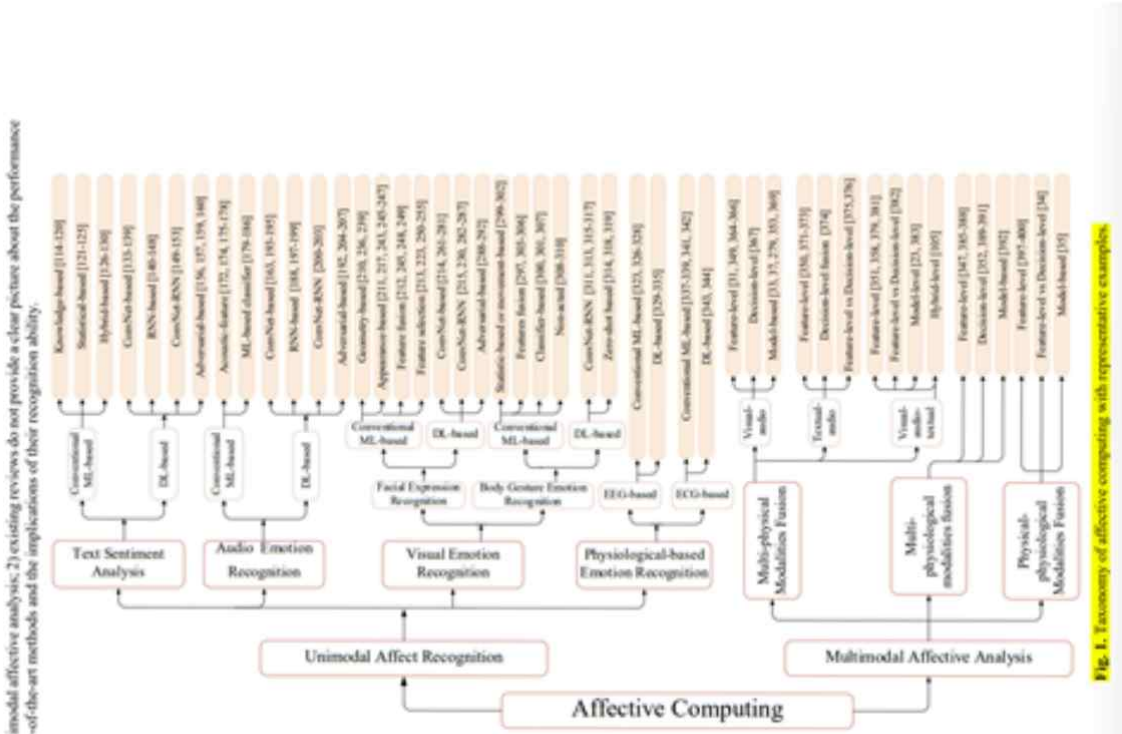
4. 연구의 기여도

이 논문은 다음과 같은 주요 기여점을 가지고 있습니다.

- **새로운 분류 체계**: 감성 컴퓨팅을 **단일 모달 감정 인식**과 **멀티모달 감성 분석**으로 분류하고, 데이터 모달리티에 따라 세분화한 최초의 논문입니다.
- **광범위한 문헌 검토**: 지난 15년간 출판된 350편 이상의 주요 논문을 체계적으로 검토하여 복잡한 연구 영역을 쉽게 탐색할 수 있는 분류 체계를 제공합니다.
- **최신 방법론 분석**: 기존의 기계 학습 기반 방법과 최신 딥러닝 기술을 모두 고려한 포괄적인 분석을 제시합니다.
- **데이터베이스 및 성능 비교**: 벤치마크 데이터베이스를 유형별로 분류하고, 주요 방법론의 특성과 정량적 성능을 비교하여 요약합니다.
- **미래 방향 제시**: 감성 컴퓨팅의 현재 어려움과 잠재적 요인을 논의하여 향후 연구에 대한 방향을 제시합니다.

<그림 1> 감성컴퓨팅 분류체계

(1-1) 감성컴퓨팅 분류체계



이 그림은 '감성 컴퓨팅'의 분류 체계를 보여주는 계층적 다이어그램입니다.

감성 컴퓨팅은 크게 두 가지 주요 접근법인 '단일 모드 감정 인식(Unimodal Affect Recognition)'과 '다중 모드 감정 분석(Multimodal

Affect Analysis)'으로 나뉩니다.

1. 단일 모드 감정 인식 (Unimodal Affect Recognition)

단일 모드 감정 인식은 한 가지 종류의 데이터(텍스트, 오디오, 시각, 생리적 신호)만을 사용하여 감정을 인식하는 방법입니다.

- **텍스트 감정 분석 (Text Sentiment Analysis)**
  - **전통적 머신러닝(Conventional ML-based):** 텍스트를 기반으로 감정(긍정, 부정, 중립 등)을 분류하는 전통적인 방법입니다.
    - **예시:** 영화 리뷰 "이 영화 정말 재밌어요!"를 **긍정**으로, "기대 이하였어요."를 **부정**으로 분류하는 것.
  - **딥러닝(DL-based):** 딥러닝 기술을 활용하여 텍스트의 감정을 분석하는 방법입니다.
    - **예시:** 고객 서비스 챗봇이 사용자의 메시지에서 분노, 불만족과 같은 감정을 파악하여 상담원에게 연결하는 것.

- **오디오 감정 인식 (Audio Emotion Recognition)**
  - **전통적 머신러닝(Conventional ML-based):** 오디오 신호의 특징(음조, 억양 등)을 추출하여 감정을 인식합니다.
    - **예시:** 콜센터에서 고객의 목소리 톤이 높아지는 것을 감지하여 불만 상태임을 파악하는 것.
  - **딥러닝(DL-based):** 딥러닝 기반으로 오디오에서 감정을 인식하는 방법입니다.
    - **예시:** 가상 비서가 "좋아!"라고 말하는 목소리의 흥분을 감지하여 사용자가 즐거워하고 있음을 아는 것.

- **시각 감정 인식 (Visual Emotion Recognition)**
  - **표정 인식(Facial Expression Recognition):** 얼굴 표정을 분석하여 감정을 인식합니다.
    - **전통적 머신러닝(Conventional ML-based):** 전통적인 머신러닝 모델을 사용합니다.
      - **예시:** 사진 속 인물의 눈썹, 입꼬리 등 **얼굴의 특정 지점**을 분석하여 웃고 있는지 아닌지를 판단하는 것.
    - **딥러닝(DL-based):** 딥러닝 모델을 사용합니다.
      - **예시:** 카메라가 사용자의 얼굴을 실시간으로 분석해 기쁨, 슬픔, 놀람 등 **7가지 기본 감정**을 인식하는 것.
  - **신체 제스처 인식(Body Gesture Emotion Recognition):** 신체 움직임이나 제스처를 통해 감정을 인식합니다.

        - **전통적 머신러닝(Conventional ML-based):** 전통적인 모델을 사용합니다.
            - **예시:** 손을 흔드는 제스처를 '환영' 또는 '인사'로 인식하는 것.
        - **딥러닝(DL-based):** 딥러닝 모델을 사용합니다.
            - **예시:** 발표자가 팔짱을 끼고 있는 것을 '방어적' 또는 '불안정'으로 해석하는 것.

- **생리 신호 기반 감정 인식 (Physiological-based Emotion Recognition)**
    - **EEG 기반(EEG-based):** 뇌전도(EEG) 신호를 분석하여 감정을 인식합니다.
        - **예시:** 광고 시청 중 뇌파 변화를 통해 시청자의 흥분도나 집중도를 측정하는 것.
    - **ECG 기반(ECG-based):** 심전도(ECG) 신호를 분석하여 감정을 인식합니다.
        - **예시:** 스마트 워치가 심박수 변화를 감지해 사용자가 스트레스를 받고 있는지 파악하는 것.
    - **IR 기반(IR-based):** 적외선 신호를 기반으로 감정을 인식합니다.
        - **예시:** 운전 중 운전자의 안면 온도 변화를 측정하여 졸음이나 긴장 상태를 감지하는 것.

2. 다중 모드 감정 분석 (Multimodal Affect Analysis)

다중 모드 감정 분석은 두 가지 이상의 데이터를 결합하여 더 정확하게 감정을 분석하는 방법입니다.

- **다중 양식 융합 (Multi-modalities Fusion)**
  - **시각-오디오(Visual-audio):** 시각 정보와 오디오 정보를 함께 사용하여 감정을 분석합니다.
    - **예시:** 대화하는 사람의 표정(시각)과 목소리 톤(오디오)을 종합적으로 분석하여 진정한 감정을 판단하는 것.
  - **텍스트-시각(Text-visual):** 텍스트와 시각 정보를 함께 사용하여 감정을 분석합니다.
    - **예시:** 소셜 미디어 게시물(텍스트)과 함께 올라온 사진(시각)을 분석하여 사용자의 감정 상태를 파악하는 것.
  - **시각-오디오-텍스트(Visual-audio-textual):** 시각, 오디오, 텍스트 정보를 모두 사용하여 감정을 분석합니다.
    - **예시:** 온라인 회의 중 발화자의 얼굴 표정, 목소리 톤, 그리고 발화 내용을 동시에 분석하여 감정 상태를 정확하게 이해하는 것.
- **다중 생리 신호 융합 (Multi-physiological modalities Fusion)**
  - **물리적-생리적 융합 (Physical-physiological Fusion):** 신체 움직임과 생리적 신호를 결합하여 감정을 분석합니다.
    - **예시:** 사용자의 걸음 속도(물리적)와 심박수(생리적)를 동시에 분석하여 사용자가 불안한 상태임을 판단하는 것.

결론적으로, 이 다이어그램은 감성 컴퓨팅의 다양한 하위 분야와 각각에 사용되는 대표적인 기술(전통적 머신러닝, 딥러닝, 특징 기반 모델 등)을 체계적으로 보여줍니다. 각 분류 아래에는 관련 연구 논문의 예시가 번호로 표시되어 있어, 해당 분야에 대한 심도 깊은 연구를 위한 참고 자료 역할도 하고 있습니다.

(2) 관련연구

2. 감성 컴퓨팅 관련 연구 요약

이 논문은 감성 컴퓨팅 분야의 기존 연구들을 세 가지 주요 관점에서 정리하고 있습니다.

1. 물리적 데이터 기반 감정 인식

설명: 사람의 감정을 텍스트, 음성, 얼굴 표정, 몸짓과 같이 외부로 드러나는 신체적 정보(물리적 데이터)를 활용해 인식하는 연구들을 말합니다.

- **주요 연구 동향:**
  - **얼굴 표정 인식(FER)**: 주로 얼굴의 미세한 표정 변화를 분석하는 연구들이 많이 이루어졌습니다.
  - **음성 감정 인식(SER)**: 목소리의 톤이나 억양을 분석해 감정을 파악하는 연구가 진행되었습니다.
  - **텍스트 감성 분석(TSA)**: 글에 담긴 긍정, 부정, 중립적인 감정이나 의견을 분석하는 연구들이 있습니다.
- **특징:** 이러한 연구들은 **딥러닝(DL) 기반**의 방법론을 많이 사용하며, 최근에는 사람의 대화 맥락을 고려하거나 여러 가지 신체적 신호를 종합적으로 분석하는 연구로 발전하고 있습니다.

2. 생리적 신호 기반 감정 인식

설명: 뇌파(EEG), 심전도(ECG)와 같이 개인이 의도적으로 숨기기 어려운 몸속의 생체 신호를 활용해 감정을 파악하는 연구입니다. 이는 "사회적 가장"으로 인해 감정을 숨기는 경우에도 정확한 감정 인식이 가능하다는 장점이 있습니다.

- **주요 연구 동향:**
  - **EEG 기반 감정 인식**: 뇌파 신호를 분석하여 감정 상태를 알아내는 연구가 활발히 진행되었습니다.
  - **ECG 기반 감정 인식**: 심전도 신호의 변화를 분석하여 감정을 예측하는 연구가 있습니다.
- **특징:** 초반에는 전통적인 기계 학습(ML)을 주로 사용했지만, 최근에는 **딥러닝** 모델을 적용하여 정확도를 높이는 연구들이 대두

되고 있습니다.

3. 물리-생리적 융합 분석

설명: 가장 최신의 연구 동향으로, 텍스트, 음성, 시각과 같은 물리적 데이터와 EEG, ECG와 같은 생리적 신호를 결합하여 감정을 분석하는 방법입니다. 인간의 감정은 다양한 방식으로 표현되기 때문에 여러 종류의 데이터를 함께 사용하면 단일 데이터만 사용하는 것보다 훨씬 더 정확하게 감정을 파악할 수 있습니다.

- **주요 연구 동향:**
  - **데이터 융합**: 오디오-시각, 텍스트-오디오, 오디오-시각-텍스트 등 다양한 물리적 데이터를 융합하는 연구가 이루어졌습니다.
  - **물리적-생리적 융합**: 특히, 음성-시각 정보와 생리적 신호를 결합하여 실시간 감정 상태를 파악하는 연구가 주목받고 있습니다.
- **특징:** 여러 데이터를 효과적으로 결합하기 위한 **다양한 융합 전략**들이 연구되고 있으며, 이러한 멀티모달 융합 방식은 단일 모달 분석 시스템보다 더 뛰어난 성능을 보입니다.

결론적으로, 이 논문은 기존 연구들이 주로 특정 분야에만 초점을 맞췄던 한계에서 벗어나, 물리적, 생리적, 그리고 이 둘을 융합하는 다양한 감성 컴퓨팅 연구들을 종합적으로 정리하고 있습니다. 이를 통해 이 복잡한 분야의 전체적인 흐름을 한눈에 파악할 수 있도록 돕는 것이 이 논문의 중요한 기여점입니다.

(3) 감정모델

3. 감성 모델 설명

감성 모델은 인간의 감정을 체계적으로 분류하고 설명하기 위해 사용되는 이론적 틀입니다. 감정을 개별적인 범주로 나누는 **이산 감정 모델**과 감정을 연속적인 차원으로 분석하는 **차원적 감정 모델**로 크게 나눌 수 있습니다.

1. 이산 감정 모델

이산 감정 모델(Discrete Emotion Model)은 감정이 독립적이고 분리된 범주로 존재한다고 가정합니다. 가장 잘 알려진 예시 중 하나는 **플루칙의 감정 바퀴(Plutchik's Wheel Model)**입니다.

- **기본 감정:** 플루칙은 기쁨, 슬픔, 신뢰, 혐오, 두려움, 분노, 놀람, 기대를 8가지 기본적인 감정으로 제안했습니다.
- **강도와 관계:** 이 모델은 이모티콘과 같은 단순한 감정 표현부터 복합적인 감정까지 설명합니다. 감정 바퀴에서 중심에서 멀어질수록 감정의 강도가 강해지며, 바퀴의 반대편에 위치한 감정은 서로 상반된 관계를 가집니다. **예를 들어, '기쁨'의 반대편에 '슬픔'이 위치하여 두 감정이 서로 상반됨을 나타냅니다.**

2. 차원적 감정 모델

차원적 감정 모델(Dimensional Emotion Model)은 감정을 2~3개의 연속적인 축(차원)으로 설명합니다. 이 접근 방식은 이산 모델이 포착하기 어려운 복합적이고 미묘한 감정의 차이를 표현하는 데 유용합니다.

2.1. Pleasure-Arousal-Dominance (PAD) 모델

PAD 모델은 감정을 세 가지 주요 차원으로 설명합니다.

- **Pleasure (쾌락):** 감정의 긍정성 또는 부정성(긍정적-부정적)을 나타냅니다. **Valence(가치)**라고도 불리며, '행복'은 높은 쾌락, '슬픔'은 낮은 쾌락에 속합니다. **예시: 즐거운 소식을 들었을 때 느끼는 '기쁨'은 높은 쾌락에 속합니다.**
- **Arousal (각성):** 감정의 에너지 수준(활성화-비활성화)을 나타냅니다. '흥분'은 높은 각성, '평온'은 낮은 각성에 속합니다. **예시: 무서운 영화를 볼 때 심장이 두근거리는 '두려움'은 높은 각성 상태입니다.**
- **Dominance (지배):** 감정을 통제하고 주도하는 느낌(지배적-복종적)을 나타냅니다. '자신감'은 높은 지배, '두려움'은 낮은 지배에 속합니다. **예시: 발표를 성공적으로 마쳤을 때 느끼는 '자신감'은 높은 지배감에 해당합니다.**

2.2. Valence-Arousal (V-A) 모델

V-A 모델은 PAD 모델에서 **지배(Dominance)** 차원을 제외한 **가치(Valence)**와 **각성(Arousal)** 두 가지 차원으로 감정을 설명합니다.

- **가치(Valence) 축:** 긍정성(즐거움)에서 부정성(불쾌함)까지의 연속체를 나타냅니다.
- **각성(Arousal) 축:** 활성화(긴장, 흥분)에서 비활성화(평온, 이완)까지의 연속체를 나타냅니다.

이 두 축을 사용하여 감정은 4개의 사분면에 위치할 수 있습니다.

- Positive Valence, High Arousal (긍정적-높은 각성): 행복, 흥분, 기쁨
  예시: 놀이기구를 탈 때 느끼는 짜릿함과 즐거움
- Negative Valence, High Arousal (부정적-높은 각성): 분노, 두려움, 스트레스
  예시: 시험을 망쳤을 때 느끼는 분노나 발표 직전의 긴장감
- Negative Valence, Low Arousal (부정적-낮은 각성): 슬픔, 우울, 피곤함
  예시: 비 오는 날 혼자 있을 때 느끼는 외로움
- Positive Valence, Low Arousal (긍정적-낮은 각성): 평온, 만족, 이완
  예시: 따뜻한 햇살 아래서 낮잠을 잘 때 느끼는 편안함

4. 감성 컴퓨팅 데이터베이스 🧠

감성 컴퓨팅(Affective Computing)은 인간의 감정을 인식, 해석, 처리, 그리고 시뮬레이션하는 시스템 및 장치 개발을 다루는 학문입니다. 이러한 기술의 핵심은 정확한 감정 분석을 위한 풍부하고 다양한 데이터베이스에 있습니다. 감성 컴퓨팅은 텍스트, 음성, 시각, 생리적 신호 등 다양한 양식(modality)의 데이터를 활용하여 감정을 파악하며, 각 데이터 유형에 따라 특화된 데이터베이스가 존재합니다.

1. 텍스트 데이터베이스

텍스트 기반 감정 분석 데이터베이스는 단어, 문장, 또는 문서에 감정적 태그가 부착되어 있습니다. 이는 주로 사용자 후기, 리뷰, 소셜 미디어 게시물 등에서 감정을 분류하는 데 활용됩니다.

- **Multi-domain sentiment (MDS):** Amazon.com의 10만 개 이상의 상품 후기 문장으로 구성되어 있으며, 긍정/부정의 이진 분류와 5단계 세부 감정(매우 긍정, 긍정, 중립, 부정, 매우 부정)으로 분류됩니다. 다양한 상품 카테고리를 포함하고 있어 일반적인 감정 분석 모델의 학습에 유용합니다.
- **IMDB:** 영화 리뷰에 특화된 대규모 데이터베이스로, 훈련 및 테스트용으로 각각 25,000개의 긍정/부정 리뷰를 제공합니다. 영화라는 단일 도메인에 집중하여 특정 분야의 감정 분석 모델을 구축하는 데 적합합니다.
- **Stanford sentiment treebank (SST):** 스탠포드 대학에서 주석을 단 의미론적 어휘 데이터베이스로, 215,154개의 구문에 세분화된 감정 라벨이 부착되어 있습니다. 문장 전체가 아닌 구문(phrases) 단위로 감정을 분류하여 더 미묘하고 복잡한 감정 분석이 가능합니다.

2. 음성/오디오 데이터베이스

음성 데이터베이스는 음성 신호의 억양, 속도, 음높이, 음색 등 비언어적 특징을 기반으로 감정을 분석합니다. 이는 크게 '비자연발생

적(simulated and induced)'과 '자연발생적(spontaneous)'으로 나눌 수 있습니다.

- **비자연발생적(Simulated and Induced):**
  - **Berlin Database of Emotional Speech (Emo-DB):** 전문 성우가 연기한 500여 개의 발화문을 담고 있어 명확하고 표준화된 감정 표현을 제공합니다. 이는 감정 분류 모델의 초기 학습에 효과적이지만, 실제 대화의 복잡성을 반영하지 못할 수 있습니다.
  - **Belfast Induced Natural Emotion (Belfast):** 피험자들에게 특정 상황을 유도하여 녹음한 데이터베이스로, Emo-DB보다 더 실제에 가까운 감정을 포착할 수 있습니다. 유도된 감정은 실제 감정과 유사한 생리적 반응을 수반하기 때문에 더 신뢰성이 높습니다.
- **자연발생적(Spontaneous):** 이 유형의 데이터베이스는 통제되지 않은 환경에서 자연스럽게 발생하는 감정을 기록하여 실제 상황에 적용하기에 용이합니다. 예를 들어, 영화나 TV 쇼의 대화, 실제 고객 서비스 통화 등이 여기에 해당합니다.

3. 시각 데이터베이스

시각 데이터베이스는 얼굴 표정, 몸짓, 제스처 등 시각적 신호를 통해 감정을 분석합니다. 이는 텍스트나 음성보다 감정을 더 직접적으로 드러내므로 감정 인식의 중요한 요소로 활용됩니다.

- **얼굴 표정 데이터베이스:**
  - **JAFFE:** 10명의 일본인 여성 배우가 7가지 기본 감정(행복, 슬픔, 분노, 놀람, 공포, 혐오, 중립)을 연출한 213장의 이미지를 포함합니다. 연구 초기 단계에 널리 사용되었던 데이터셋입니다.
  - **Oulu-CASIA NIR-VIS:** 2,880장의 이미지 시퀀스를 포함하며, 적외선(NIR) 및 가시광선(VIS) 스펙트럼 데이터를 동시에 제공하여 조명 변화에 강한 모델 개발에 유용합니다.
  - **4DFAB:** 180명의 피험자가 4가지 세션에서 연기한 약 180만 개의 동적 고해상도 3D 얼굴 표정 데이터를 포함하여, 단순 이미지보다 훨씬 풍부한 3차원 감정 정보를 제공합니다.
  - **EmotioNet:** 백만 개 이상의 얼굴 이미지를 포함하며, 대규모의 다양한 데이터셋을 통해 감정 인식 모델의 성능을 향상시키는 데 기여합니다.

- **몸짓 감정 데이터베이스:**
  - **FAce and BOdy database (FABO):** 얼굴과 몸짓을 모두 포함하는 최초의 공개적인 바이모달(bimodal) 데이터베이스입니다. 감정을 표현하는 데 얼굴뿐만 아니라 몸 전체의 움직임이 중요하다는 점을 반영합니다.
  - **Emotional body expression in daily actions database (EMILYA):** 일상적인 동작 속에서 몸짓을 수집한 데이터베이스로, 전신 동작의 3D 데이터도 포함하여 자연스러운 상황에서의 감정 분석을 가능하게 합니다.

4. 생리적 데이터베이스

생리적 데이터는 의식적인 통제가 어렵기 때문에 텍스트, 음성, 시각 데이터보다 감정을 더 객관적이고 신뢰성 있게 나타내는 장점이 있습니다. 심박수, 피부 전도율(GSR), 뇌파(EEG), 근전도(EMG) 등이 주로 사용됩니다.

- **DEAP:** 32명의 피험자가 40개의 뮤직비디오를 보면서 측정한 32채널 EEG, EOG, EMG, RESP, GSR 및 체온 데이터를 포함합니다. 음악을 통한 감정 유발 실험에 적합합니다.
- **Detecting Stress during Real-World Driving Tasks (DsDrD):** 운전 중 운전자의 스트레스 수준을 측정하는 데 사용되는 데이터베이스로, 실제 상황에서 발생하는 감정을 분석하는 데 유용합니다.
- **Wearable Stress and Affect Detection (WESAD):** 스트레스 감지를 위해 제작된 멀티모달 고품질 데이터베이스입니다. 착용형 기기(wearable device)에서 수집된 데이터를 포함하여 실제 사용자 환경에서의 감정 분석 연구에 활용됩니다.

5. 멀티모달 데이터베이스

인간은 감정을 표현할 때 여러 가지 채널(모달리티)을 동시에 사용합니다. 따라서 텍스트, 음성, 시각 등 여러 데이터를 결합한 멀티모달 데이터베이스는 더 정확하고 풍부한 감정 분석을 가능하게 합니다.

- **IEMOCAP:** 전문 성우의 대본 또는 즉흥적인 상황에서 얻은 감정 데이터를 담고 있으며, 얼굴, 손, 머리 등의 세부적인 정보가 포함되어 있어 음성, 표정 등 다양한 감정 신호를 종합적으로 분석할 수 있습니다.
- **CMU-MOSI:** 감정 및 정서 인식을 위한 가장 큰 멀티모달 데이터베이스 중 하나입니다. 주로 비디오 블로그 영상 데이터를 기반으로 텍스트, 음성, 시각적 특성을 종합적으로 분석하여 감정을 예측합니다.

- **RECOLA:** 원격 화상 회의에서 재난 탈출 계획을 논의하는 상황에서 자발적으로 발생한 멀티모달 데이터를 포함합니다. 통제되지 않은 실제 상황에서의 감정 분석 모델을 개발하는 데 매우 중요한 자원입니다.

(5-2) 음성 감성인식

5.2 음성 감정 인식(SER)

SER은 컴퓨터가 사람의 **목소리**에서 감정을 파악하는 기술입니다. 음성 비서나 콜센터 자동 응답 시스템 등에 활용될 수 있어요. 이 기술은 크게 **특징 추출**과 **분류**의 두 단계로 이루어집니다.

5.2.1 전통적인 ML 기반 SER

전통적인 방식은 목소리에서 감정을 잘 드러내는 **특징들**을 수동으로 찾아내고, 이를 바탕으로 감정을 분류합니다. 마치 전문 성우나 음악가가 음성을 분석하는 과정과 유사합니다.

- **음향 특징 추출:** 사람의 목소리에는 감정을 나타내는 여러 정보가 담겨 있어요.
  - **운율적 특징(Prosodic features):** 목소리의 **높낮이, 크기, 속도**와 같이 말의 리듬과 억양에 관련된 특징입니다.
  - **음질 특징(Voice quality features):** 목소리의 떨림(jitter), 흔들림(shimmer) 등 성대의 미세한 움직임에서 나오는 특징입니다.
  - **스펙트럼 특징(Spectral features):** 소리의 주파수 대역별 에너지 분포를 분석한 특징입니다.

  예시: "와, **정말** 대단한데요!"라고 말할 때, '정말'이라는 단어를 **높은 톤과 큰 소리**로 말하면 감탄(긍정)을 나타내고, **낮고 작은 소리**로 말하면 실망(부정)을 나타내는 것처럼, 이러한 특징들을 컴퓨터가 추출합니다.
- **분류기 선택:** 추출된 특징들을 바탕으로 감정을 예측하는 모델을 사용합니다. SVM, GMM, HMM, RF 등 다양한 모델이 사용되죠.

5.2.2 딥러닝(DL) 기반 SER

딥러닝은 마치 사람이 직관적으로 감정을 느끼듯, 목소리 전체를 듣고 복잡한 패턴을 스스로 파악합니다. 수동적인 특징 추출 과정이 필요 없어 더 효율적입니다.

- **딥 컨브넷(ConvNet) 학습:** 음성 신호를 '스펙트로그램(spectrogram)'이라는 이미지 형태로 변환하여 CNN(합성곱 신경망)으로 분석합니다. 마치 소리를 눈으로 보는 것처럼, 감정에 따라 스펙트로그램의 패턴이 어떻게 달라지는지를 학습하는 거죠.
    - **예시:** 화난 목소리 스펙트로그램은 특정 주파수 대역에서 에너지가 높게 나타나는 패턴을 보이는데, 딥러닝 모델이 이러한 패턴을 자동으로 감지합니다.
- **딥 RNN 학습:** 긴 음성 데이터의 시간적 흐름을 이해하는 데 특화된 RNN(순환 신경망)을 활용합니다. 특히 **Bi-LSTM**은 이전과 이후의 맥락을 모두 고려하여 감정을 파악합니다. **어텐션(Attention) 메커니즘**을 추가하면 문장 중 감정이 가장 강하게 드러나는 부분에 집중하여 정확도를 높입니다.
    - **예시:** "아니, **진짜**?"라는 문장에서, '진짜'라는 단어를 강조하는 부분이 분노를 나타내는 핵심적인 정보임을 어텐션 모델이 스스로 찾아냅니다.
- **딥 컨브넷-RNN 학습:** CNN으로 음성의 주파수 특징을 파악하고, RNN으로 시간적 변화를 분석하여 두 모델의 장점을 모두 활용하는 방식입니다.
- **딥 적대적 학습:** 학습 데이터가 부족하거나 특정 환경(예: 시끄러운 환경)에서 녹음된 음성 데이터가 필요할 때 유용합니다. 가상의 데이터를 만들어 훈련시키거나, 환경의 차이를 극복하도록 모델을 훈련시킵니다.
    - **예시:** GANs(생성적 적대 신경망)를 사용해 슬픈 목소리를 흉내 낸 가짜 음성 데이터를 대량으로 생성하여, 모델이 '슬픔' 감정을 더 잘 인식하도록 훈련시킵니다.

(5.3) 얼굴 표정인식+그림 4

5. 3시각적 감정 인식

시각적 감정 인식은 크게 **얼굴 표정 인식(FER)**과 **몸짓 또는 제스처 인식(EBGR)**으로 분류됩니다. 이 두 방법은 전처리, 특징 추출 및 분류 방법에서 큰 차이를 보입니다.

5.3.1 얼굴 표정 인식 (FER)

FER은 얼굴에 나타난 감정적 단서를 이미지나 비디오를 통해 분석합니다. 분석 대상에 따라 정지 이미지 기반 FER과 동적 비디오 기반 FER로 나뉘며, 표현의 강도에 따라 미세 표정(micro-FER)과 일반 표정(macro-FER)으로 구분되기도 합니다. 또한, 얼굴 이미지의 차원에 따라 2D-FER과 3D/4D-FER로 나뉩니다.

주요 FER 방법론은 특징을 직접 설계하는 **머신러닝(ML) 기반**과 자동으로 특징을 학습하는 **딥러닝(DL) 기반**으로 분류됩니다.

- **5.3.1.1 기존 머신러닝 기반 FER:**
  - 얼굴의 모양을 분석하는 **기하학적 특징(Geometry-based)**, 얼굴의 질감을 분석하는 **외형적 특징(Appearance-based)**, 그리고 이 둘을 결합한 **특징 융합(Feature fusion)**에 의존합니다.
  - 과도한 특징을 줄이기 위해 저차원 매핑, 가중치 부여, 관심 영역(ROI) 기반 선택 등의 **특징 선택** 기술이 사용됩니다.
- **5.3.1.2 딥러닝 기반 FER:**
  - 수작업으로 특징을 설계할 필요 없이 엔드투엔드(end-to-end) 방식으로 특징을 표현합니다.
  - 주로 **심층 컨볼루션 네트워크(ConvNet)**, 시간적 정보를 고려하는 **ConvNet-순환 신경망(RNN)**, 그리고 포즈나 신원 변화에 강건한 **적대적 학습(Adversarial learning, GANs)** 모델이 사용됩니다.
  - 얼굴 이미지의 중요한 부분을 강조하기 위해 다양한 **어텐션(attention) 메커니즘**이 도입되기도 합니다.

5.3.2 몸짓 감정 인식 (EBGR)

EBGR은 얼굴 표정을 정확히 포착하기 어려운 환경에서 사용되는 대안적 방법입니다. 몸 전체의 자세나 움직임, 손동작, 머리 위치와 같은 상체 정보를 활용합니다. 사람 탐지(human detection)를 전처리 과정으로 사용하며, 마찬가지로 ML 기반과 DL 기반으로 나뉩니다.

- **5.3.2.1 기존 머신러닝 기반 EBGR:**
  - 주로 신체 움직임의 통계적 특성을 분석하는 **통계 기반(Statistic-based)** 방법과 역동적인 움직임의 질을 분석하는 **동작 기반(Movement-based)** 방법으로 나뉩니다.
  - 다양한 신체 자세 특징을 결합하는 **특징 융합**을 통해 인식 성능과 강건성을 향상시킵니다.
  - 결정 트리, SVM, 신경망 등 다양한 **분류기(Classifier)**를 사용합니다.
  - 배우의 과장된 연기가 아닌 자연스러운(non-acted) 감정 인식에 대한 연구가 이루어지고 있습니다.
- **5.3.2.2 딥러닝 기반 EBGR:**
  - 엔드투엔드 방식의 감정 인식을 위해 **ConvNet, LSTM, ConvNet-LSTM** 등의 네트워크 구조를 사용합니다.
  - 주어진 범주에 없는 새로운 감정을 인식하기 위해 **제로샷(Zero-shot) 기반 학습** 프레임워크가 도입되었습니다.

(5.4) 생체신호이용 감정 인식과정+그림 5

5. 4 생체 신호를 이용한 감정 인식 과정

그림 5는 생체 신호 기반 감정 인식 시스템이 어떻게 작동하는지 5단계로 보여줍니다.

1. **자극(Stimulating)**: 이 단계는 감정을 유발하기 위한 자극을 제공하는 단계입니다. 그림에서는 이미지, 음악, 비디오를 예시로 들고 있습니다. 이러한 자극을 통해 사람의 감정 상태를 변화시킵니다.
2. **신호 기록 및 추출(Recording & Extracting)**: 자극에 반응하여 나타나는 생체 신호를 기록합니다. 주요 신호로는 뇌파(EEG), 피부 전도 반응(RESP), 심박수(HR), 근전도(EMG), 심전도(ECG) 등이 있습니다. 이 단계에서는 기록된 신호에서 감정 인식을 위한 중요한 특징들을 추출합니다.
3. **특징 처리(Features Processing)**: 추출된 생체 신호의 특징을 처리하는 단계입니다. 이 과정은 신호 전처리, 특징 분석, 특징 선택 및 축소 등을 포함하며, 분류 모델에 적합한 데이터 형태로 가공합니다.
4. **학습 및 분류(Training & Classifying)**: 처리된 특징 데이터를 사용하여 감정 분류 모델을 학습시킵니다. 그림에서는 SVM, KNN, LDA, RF, NB, NN과 같은 다양한 분류 모델을 예시로 보여줍니다. 이렇게 학습된 모델은 새로운 데이터가 들어왔을 때 해당 감정을 예측할 수 있게 됩니다.
5. **감정 인식(Emotion Recognition)**: 마지막 단계에서는 학습된 모델을 통해 감정을 최종적으로 인식합니다. 감정은 이산적 감정 모델(예: 행복, 슬픔, 분노) 또는 차원적 감정 모델(예: 쾌락-각성)로 분류될 수 있습니다.

이 그림은 생체 신호를 활용하여 감정을 객관적이고 신뢰성 있게 인식하는 방법의 전반적인 흐름을 잘 보여주고 있습니다. 특히 뇌파(EEG)와 심전도(ECG)는 이러한 접근 방식에 가장 많이 사용되는 신호로 언급됩니다.

6. 멀티모달 감정 분석

- 멀티모달 감정 분석 (Multimodal affective analysis):
  하나의 데이터(음성, 시각 등)만을 사용하는 '단일 모드'와 달리, 여러 종류의 데이터를 융합하여 사람의 감정을 분석하는 기술입니다. 이를 통해 더 정확하고 포괄적인 이해를 얻을 수 있습니다.
- 데이터 융합 방식 (Fusion Strategies):
  다양한 데이터를 결합하는 방식에 따라 크게 네 가지 전략으로 나눌 수 있습니다.

6. 특징 레벨 융합 (Feature-level fusion):
   다양한 모드에서 얻은 특징들을 하나의 큰 특징 벡터로 결합한 후, 분류기에 입력하여 분석합니다.
7. 결정 레벨 융합 (Decision-level fusion):
   각 모드를 독립적으로 분석하여 각각의 감정 예측 결과를 얻은 후, 이 결과들을 종합하여 최종 결론을 내립니다.
8. 모델 레벨 융합 (Model-level fusion):
   여러 모델을 결합하여 데이터 간의 상호 관계를 파악하고, 이를 통해 감정을 분석합니다.
9. 하이브리드 융합 (Hybrid fusion):
   위에서 설명한 두 가지 이상의 융합 방식을 조합하여 사용합니다.

그림 6. 여러 가지 융합 전략을 사용한 주요 사례들을 보여줍니다.

10. 특징 레벨 융합:
    다중 모드 입력에서 특징을 추출하여 하나의 일반적인 특징 벡터를 형성하고, 이것이 분류기로 보내집니다. **그림 6(a), (b), (c)**는 시각-오디오, 시각-텍스트, 오디오-텍스트와 같은 다중 물리적 모드 간의 특징 레벨 융합을 기반으로 한 몇 가지 예시를

보여줍니다.

11. 결정 레벨 융합:
    각 모드에서 독립적으로 생성된 여러 결정 벡터를 하나의 특징 벡터로 연결합니다. **그림 6(d)**는 EGG, ECG 및 EDA의 다중 생리적 모드에 대한 결정 레벨 융합을 기반으로 한 예시를 보여줍니다.
12. 모델 레벨 융합:
    다양한 모드에서 추출된 특징 간의 상관 관계를 발견하고, HMM 및 2단계 ELM [353]과 같은 완화되고 부드러운 유형으로 융합 모델을 사용하거나 설계합니다. **그림 6(e) 및 (f)**는 모델 레벨 융합의 두 가지 예시입니다. 전자는 생리적-물리적 모달리티를 위한 것이고, 후자는 시각-오디오-텍스트 모달리티를 위한 것입니다.
13. 하이브리드 융합:
    특징 레벨 융합과 결정 레벨 융합을 모두 조합한 방식입니다. **그림 6(g)**는 시각-오디오-텍스트 모달리티를 위한 하이브리드 융합 프레임워크를 보여줍니다.

6.1 물리적 모달리티 융합을 통한 감정 분석

인기 있는 모달리티 조합 방식을 고려하여, 멀티모달 감정 분석을 위한 **물리적 모달리티 융합**을 시각-오디오 감정 인식 [354, 31], 텍스트-오디오 감정 인식 [355, 356], 그리고 시각-오디오-텍스트 감정 인식 [357, 358]으로 분류할 수 있습니다. 다음은 각 방식에 대한 대표적인 방법들을 요약한 것입니다.

6.1.1 시각-오디오 감정 인식

시각 및 오디오 신호는 일상생활에서 사람들이 소통할 때 감정을 보여주는 가장 자연스럽고 효과적인 감정 신호입니다 [359]. 많은 연구 [360-363, 354]에서 시각-오디오 감정 인식이 단일 모드인 시각 또는 오디오 감정 인식보다 더 뛰어난 성능을 보인다고 합니다.

- 특징 레벨 융합:

Chen et al. [349]은 ML 기반 시각-오디오 감정 인식을 제안했습니다. 이 방법은 동적 HOG-TOP 텍스처 특징과 음향 특징 또는 기하학적 특징을 융합하여 사용합니다. Tzirakis et al. [31]은 CNN과 심층 잔차 네트워크를 사용하여 각각 오디오 및 시각적 특징을 추출한 다음, 이들을 연결하여 LSTM에 입력해 감정 값을 예측했습니다. 다양한 어텐션 메커니즘 또한 시각-오디오 감정 인식 작업에 성공적으로 적용되었습니다 [364, 365].

- 결정 레벨 융합:
  Hao et al. [367]은 다중 작업 및 블렌딩 학습을 기반으로 하는 앙상블 시각-오디오 감정 인식 프레임워크를 제안했습니다. 이 방법은 오디오 및 시각 신호의 두 가지 핸드크래프트(Handcraft) 기반 및 DL(Deep Learning) 기반 특징으로 구성된 네 가지 하위 모델을 블렌딩 앙상블 알고리즘을 기반으로 융합하여 최종 감정을 예측합니다.
- 모델 레벨 융합:
  이 방식은 ML 기반 모델(예: HMM, Kalman filters 및 DBN)이 융합된 정보를 학습하고 동시에 결정을 내리도록 요구합니다. Lin et al. [33]은 오디오-시각 신호의 시간적 관계를 정렬하기 위해 SC-HMM(semi-coupled HMM)을 제안했으며, Glodek et al. [369]은 마르코프 모델을 기반으로 한 Kalman 필터를 설계하여 시간적으로 정렬된 분류기 결정을 결합하여 감정 상태를 인식합니다. Zhang et al. [37]은 CNN을 통해 오디오-시각 신호에서 특징을 추출하고 DBN을 통해 통합한 후, 선형 SVM 분류기를 사용하여 감정을 인식했습니다.

6.3 물리적-생리적 모달리티 융합을 통한 감정 분석

감정의 변화는 복잡한 심리-생리적 활동과 관련이 있습니다. 따라서 연구자들은 물리적 신호와 생리적 신호를 동시에 활용하여 감정을 분석하는 데 집중해 왔습니다. 그러나 비디오와 생리적 감정 데이터베이스가 많지 않기 때문에 관련 연구는 아직 적은 편입니다.

- 특징 레벨 융합:
  이 방법은 **생체 신호(EEG, ECG 등)**와 **물리적 신호(비디오, 음악)**를 함께 사용하여 감정 인식 모델을 만듭니다. 예를 들어, LSTM을 사용해 뇌파(EEG)와 비디오에서 얻은 얼굴 특징을 융합하여 감정을 예측하는 방식입니다. 또 다른 연구에서는 EDA(피

부전도 반응)와 음악을 융합하여 감정을 인식하는 방법을 제안했습니다.

- 결정 레벨 융합:

  Huang et al. [34]은 비디오-뇌파(EEG) 기반 멀티모달 감정 분석을 제안했습니다. 이 방법은 비디오의 얼굴 표정과 뇌파의 스펙트럼 전력 데이터를 특징 레벨과 결정 레벨 관점에서 융합하여 감정을 분석합니다. 연구 결과에 따르면, 멀티모달 감정 분석은 단일 모드 감정 인식보다 성능이 우수하며, 융합 전략 중에서는 결정 레벨 융합이 특징 레벨 융합보다 더 나은 성능을 보였다고 합니다.

- 모델 레벨 융합:

  Wang et al. [35]은 **깊은 멀티모달 심층 신념 네트워크(deep multimodal DBN)**를 설계하여 다양한 심리-생리적 신호와 비디오 신호를 융합하고 최적화했습니다. 모든 모달리티의 특징이 통합된 후, SVM을 사용해 감정을 인식하는 데 활용되었습니다.

(6)가중모드 감정 분석+그림 6

<다중 모드 감정 분석의 분류 그림 6 설명>

제공해주신 이미지는 멀티모달(multimodal) 감정 분석에 대한 다양한 접근 방식을 분류한 그림입니다. '멀티모달'이란 오디오, 비디오, 텍스트 등 여러 종류의 데이터를 동시에 사용하여 감정을 인식하는 것을 의미합니다. 그림 6은 크게 세 가지 주요 감정 융합(fusion) 방법으로 나뉩니다.

1. 특징 수준 융합 (Feature-level Fusion)

이 방법은 각 모드(오디오, 비디오, 텍스트)에서 추출한 특징들을 먼저 결합한 후, 결합된 특징을 기반으로 감정을 인식합니다.

쉬운 예시:
여러 사람이 한 사람의 감정을 추리한다고 상상해 보세요.

- **A:** "음, 목소리 톤이 밝네." (오디오 특징)
- **B:** "표정이 웃고 있네." (비디오 특징)
- C: "방금 '정말 기뻐!'라고 말했어." (텍스트 특징)
  이 세 가지 특징을 모두 조합하여 "이 사람은 기쁘다"는 결론을 내리는 것과 같습니다.
- (a) Chen et al.34
  : 오디오, 비디오, 텍스트 특징을 결합하여 감정 분류기를 학습합니다.
- (b) Poria et al.35
  : CNN과 LSTM을 사용하여 각 모드의 특징을 추출하고, 이를 결합하여 감정 분석에 활용합니다.
- (c) Mittal et al.351
  : 오디오와 비디오 특징을 결합하여 감정 분류를 수행합니다.

2. 결정 수준 융합 (Decision-level Fusion)

이 방법은 각 모드별로 독립적으로 감정 예측을 수행한 후, 마지막 단계에서 각 모드의 예측 결과를 결합하여 최종 감정을 결정합니다.

쉬운 예시:
이번에는 각자 독립적으로 추리한 결과를 모아서 최종 결정을 내리는 상황입니다.

- **A:** "목소리만 들어보니 기쁜 것 같아." (오디오 예측)
- **B:** "표정만 보니 기쁜 것 같아." (비디오 예측)
- C: "말만 들어보니 기쁜 것 같아." (텍스트 예측)
  세 사람의 예측이 모두 "기쁨"으로 일치하니, "이 사람은 기쁘다"는 최종 결론을 내립니다.
- (d) Yang352
  : 오디오-비디오-생리적(physiological) 특징을 개별적으로 분석한 뒤, 각 모드별 예측 결과를 통합하여 최종 결정을 내립니다.

3. 모델 수준 융합 (Model-level Fusion)

이 방법은 각 모드별로 모델을 따로 학습시킨 후, 이 모델들을 통합하여 더 나은 성능을 달성합니다.

쉬운 예시:
한 명의 의사(특징 수준 융합)가 모든 정보를 종합하는 대신, 여러 전문의가 협업하는 상황이라고 생각하면 쉽습니다.

- **성형외과 전문의:** "표정을 분석해보니 기쁘네요."
- **이비인후과 전문의:** "목소리 톤을 분석해보니 기쁘네요."
- 언어학 전문의: "말한 내용을 분석해보니 기쁘네요."
  이처럼 각 분야 전문가(모델)의 의견을 종합하여 최종 진단(감정 분석)을 내리는 방식입니다.
- (e) Wang et al.23

: 오디오와 비디오 데이터를 개별 모델로 처리하고, 이 모델들의 출력을 결합하여 감정 인식을 수행합니다.

- (f) Akhtar et al.25
  : 오디오와 텍스트 데이터를 각기 다른 모델로 분석한 후, 이들의 결과를 통합합니다.

4. 하이브리드 수준 융합 (Hybrid-level Fusion)

하이브리드 방식은 위의 방법들을 결합하여 사용합니다.

쉬운 예시:
특징 수준 융합과 모델 수준 융합을 섞어서 사용합니다. 예를 들어, 오디오와 비디오 특징을 먼저 합쳐서 분석한 뒤, 이 결과를 다른 모델과 함께 다시 분석하는 복합적인 방식입니다.

- (g) Wollmer et al.105
  : 오디오와 비디오 특징을 먼저 추출하여 융합한 뒤, LSTM 모델을 통해 감정 예측을 수행합니다.

(7) 논의 논문내용 요약

7. 논의 (Discussions)

이 논문은 감정 분석에 대한 내용을 전반적으로 다루고 있습니다. 사람의 감정을 파악하는 다양한 방법과 그 기술들을 쉽게 설명합니다.

7.1 감정 인식을 위한 다양한 신호들 (Unimodal affect recognition)

감정 인식을 위해 얼굴 표정, 목소리, 글 등 한 가지 신호만 사용하는 방식을 설명합니다. 이 중에서는 특히 얼굴 표정이나 몸짓 같은 시각적 신호가 가장 널리 쓰입니다.

7.2 여러 신호를 합쳐서 감정을 파악하는 방법 (Multimodal affective analysis)

글, 목소리, 영상을 함께 사용하는 등 여러 신호를 결합하면 한 가지만 사용할 때보다 감정을 더 정확하게 파악할 수 있습니다. 여러 데이터를 합치는 방식에 따라 분석 결과가 달라집니다.

7.3 인공지능 기술의 발전 (ML and DL methods)

예전에는 전통적인 기계 학습(ML) 기술을 사용했지만, 요즘은 인공지능이 스스로 학습하는 딥러닝(DL) 기술을 주로 씁니다. 딥러닝은 데이터를 자동으로 분석해 더 좋은 성능을 보여줍니다.

7.4 데이터의 중요성 (Effects of databases)

좋은 감정 데이터베이스가 많아질수록 감정 분석 기술도 함께 발전합니다. 데이터의 양과 질이 기술의 성패를 좌우합니다.

7.5 감정 분석 결과 평가 (Performance metrics)

감정 분석 기술의 성능을 평가할 때는 정확도, 정밀도, 재현율 같은 다양한 지표를 사용합니다.

7.6 감정 분석 기술의 활용 (Applications)

이 기술은 추천 시스템, 의료, 교육, 게임 등 우리의 실생활 곳곳에 다양하게 적용될 수 있으며, 앞으로 그 활용 범위는 더욱 넓어질 것입니다.

이 문서는 정서 컴퓨팅(Affective Computing)에 대한 최신 연구 동향과 앞으로의 발전 방향을 다루고 있습니다.

주요 내용

14. **연구 동향**: 350편 이상의 논문을 검토하여 정서 컴퓨팅의 다양한 측면(단일/다중 양식, 기존 머신러닝/딥러닝 등)을 분석했습니다.
15. **정서 모델**: 정서 모델을 심리학 이론 기반의 이산/차원 모델과 생리학적 신호 기반의 정서 모델로 분류했습니다.
16. **데이터베이스**: 텍스트, 음성, 시각, 생리학적, 그리고 다중 양식 데이터를 포함하는 공통으로 사용되는 데이터베이스를 조사했습니다.
17. **최신 접근 방식**: 딥러닝 기반의 단일 양식 및 다중 양식 접근 방식이 정서 컴퓨팅의 성능을 향상시키는 데 기여했다고 언급합니다.
18. **융합 전략**: 융합 전략은 정서 컴퓨팅의 성능을 크게 향상시킬 잠재력이 있으며, 규칙 기반, 지식 기반, 그리고 역할 기반 융합이 중요한 역할을 합니다.

향후 연구 방향

- **확장된 데이터베이스 구축**: 자발적, 비자발적 정서를 모두 포함하는 다양하고 확장된 데이터베이스를 개발해야 합니다.
- **어려운 과제 해결**: 부분 폐색이나 가짜 표정과 같은 어려운 문제를 해결하는 과제를 다뤄야 합니다.
- **신경망 아키텍처**: 이산 및 차원 정서 모델을 모두 지원하는 적합한 신경망 아키텍처를 설계하는 것이 중요합니다.
- **융합 전략 개선**: 융합 전략을 더 발전시켜야 합니다.
- **비지도 학습**: 자율 학습 방식이 정서 인식의 견고성과 안정성을 높이는 데 도움이 될 수 있습니다.

- **로봇 공학 적용**: 로봇이 인간의 정서를 모방하고 적절하게 반응하도록 만드는 데 정서 컴퓨팅 기술을 적용할 수 있습니다.

각 방향에 대한 예시

- **확장된 데이터베이스 구축**: 사람들이 일부러 꾸며낸 미소와 진심으로 웃는 미소를 구분할 수 있는 데이터셋을 만드는 것을 예로 들 수 있습니다.
- **어려운 과제 해결**: 마스크를 쓴 사람의 얼굴 표정을 분석하거나, 잡음이 심한 환경에서 음성 톤을 정확하게 파악하는 알고리즘을 개발하는 것이 여기에 해당합니다.
- **신경망 아키텍처**: 행복, 슬픔 같은 이산적인 감정은 물론, 그 감정의 강도(매우 행복함, 조금 슬픔 등)를 동시에 파악하는 신경망 모델을 설계하는 것입니다.
- **융합 전략 개선**: 얼굴 표정 데이터와 심장 박동수 같은 생체 신호 데이터를 결합하여 정서를 더 정확하게 파악하는 시스템을 만드는 것이 좋은 예입니다.
- **비지도 학습**: 특정 감정 라벨이 없는 수많은 영상 클립을 AI가 스스로 학습하여 '슬픔'에 해당하는 패턴을 찾아내는 것입니다.
- **로봇 공학 적용**: 노인 요양 로봇이 어르신의 외로움을 감지하고 위로의 말을 건네거나 부드러운 스킨십을 제공하도록 만드는 기술에 이 연구가 활용될 수 있습니다.