

# 휴먼이해 인지컴퓨팅 기술 연구 동향

## Human-Understanding Cognitive Computing Technology Research Trends

정현태 (H.T. Jeong, htjeong@etri.re.kr)

휴먼증강연구실 책임연구원

김가규 (G.G. Kim, ggkim@etri.re.kr)

휴먼증강연구실 책임연구원

노경주 (K.J. Noh, kjnoh@etri.re.kr)

휴먼증강연구실 책임연구원

임정목 (J.M. Lim, jmlim21@etri.re.kr)

휴먼증강연구실 책임연구원

임지연 (J.Y. Lim, kusses@etri.re.kr)

휴먼증강연구실 선임연구원

정승은 (S.E. Chung, schung@etri.re.kr)

휴먼증강연구실 선임연구원

### ABSTRACT

Human behavior and emotions are influenced by experiences accumulated through the past and present. To realize artificial intelligence technology that understands and sympathizes with humans, cognitive computing technology that automatically analyzes human behaviors, habits, and emotions associated with specific situations and uses past data and domain knowledge is required. In this study, we examine the latest research trends on human-understanding cognitive computing technology that recognizes human behavior and emotions, stores them as experience data, and provides services by analyzing the stored data. Further, we introduce high-quality data collection research in real-life and services for improving physical and mental health. We also review key issues essential for developing these technologies.

**KEYWORDS** 감정인식, 인공지능, 인지컴퓨팅, 행동인식, 휴먼이해

## 1. 서론

영국의 인지신경과학자 Catherine Loveday는 우리의 지난 경험들은 의식적이든 무의식적이든, 그리고 적극적으로 떠올린 것이든 생각지도 않게 떠오른 것이든, 언제나 우리의 행동에 영향을 미치기 마련이며, 무엇보다도 기억은 자신의

정체성에서 대단히 본질적인 부분이라고 서술했다[1]. 심리학자이자 신경과학자인 Lisa Feldman Barrett은 감정은 그때 상황에 처한 사람의 감각의 의미이며, 행동의 지침이며, 뇌는 과거의 경험에 기초하여 현재의 예측을 구성한다고 주장하였다[2]. 이들은 모두 과거 특정 상황에서 반응하는 한 사람의 행동과 감정을 이해할 때 그 사람

\* DOI: <https://doi.org/10.22648/ETRI.2022.J.370103>

\* 본 연구는 한국전자통신연구원 연구운영비지원사업의 일환으로 수행되었음[21ZS1100, 자율성장형 복합인공지능 원천기술 연구].



본 저작물은 공공누리 제4유형

출처표시+상업적이용금지+변경금지 조건에 따라 이용할 수 있습니다.

©2022 한국전자통신연구원

에 대한 이해와 교감이 가능하다는 점을 강조하고 있다.

따라서, 사람과 교감이 가능한 인공지능을 구현하기 위해서는 특정 상황에서 개인의 행동과 감정을 인식하고 이에 대한 정보를 축적하여 개인의 행동과 감정을 예측하는 기술 실현이 우선 시되어야 한다. 이러한 기술 구현을 위해서는 인공지능 모델에서 학습하기 위한 대용량, 고품질의 데이터가 반드시 필요하다. 또한, 데이터로부터 인간이 상황을 인지하는 데 필요한 행동, 감정 및 환경에 대한 정보를 추출하는 기술이 필요하다. 그러나, 과거 이러한 요구사항들을 동시에 만족시키는 연구는 대부분 실험실과 같은 제한된 상황에서 이루어져 실제 상황에 적용하기에는 어려웠으며, 최근에서야 실생활 데이터를 수집하여 행동과 감정을 인식하는 다양한 시도가 이루어지고 있다.

휴면이해 인지컴퓨팅 기술은 일상생활 중 벌어지는 상황과 행동, 감정을 인식하고 예측하는 기술을 의미한다. 이러한 기술은 사용자의 신체적·정신적 건강을 관리하고 추천하는 서비스에 적용될 수 있다. 본고에서는 휴면이해 인지컴퓨팅 기술 관련 연구를 편의상 일상생활 중에 수집한 데이터셋을 이용하여 행동과 감정을 예측하는 기술, 그리고 이러한 데이터와 인식 기술을 이용한 서비스 관련 연구로 나누고, 각 기술과 관련된 연구 중 최신 출판된 논문을 중심으로 기술 동향을 분석하였다.

## II. 행동 인식 기술 동향

이 장에서는 분류대상인 ‘행동’의 정의에 대해 기술하고, HAR 데이터셋 주요 특징, 행동인식 기술, 그리고 라이프로그 데이터의 이벤트 분할 기술의 연구 동향을 분석한다.

### 1. ‘행동’의 정의

본고에서는 장소, 시간의 제약을 두지 않고 스마트폰이나 스마트워치를 사용하여 일상생활 중 사용자 행동 예측을 목표로 하는 비광학식 방법에 대한 연구 동향을 조사하였다.

인간의 행동은 선행연구에서 다양하게 정의되고 있으나, 많은 연구에서 참고문헌 [3]과 [4]에서 정의한 내용을 인용하고 있다.

참고문헌 [3]에서는 행동이 갖는 의미의 정도와 소요되는 시간에 따라 인간의 행동을 분류한다. Motion은 수초 이내에 이루어지는 행동으로, 시선 또는 머리의 자세 추정 등 비교적 작은 움직임만을 탐지하는 것으로 정의하였다. Action은 수초에서 수분 정도의 시간에서 이루어지는 움직임으로 앉기, 서기, 걷기 등 인간의 기본 행동 단위가 된다. Activity는 수분에서 수시간에 걸쳐 이루어지며, 특정한 순서를 갖는 일련의 행동들로 구성된 작업을 의미한다. 예를 들면 조리, 샤워, 침구준비 등이 있다. Behavior는 수시간, 또는 수일에 걸쳐 나타나는 일상적 행동(ADL: Activities of Daily Living)을 의미하며, 개인의 생활 방식, 습관 등을 설명할 수 있는 매우 고수준의 의미로 이해되는 행동을 의미한다. Motion에서 Action, Activity, Behavior의 순서로 행동의 지속 시간이 길어지며, 지속 시간이 길어지는 만큼 단순한 움직임이 아닌 특정한 목적을 갖는 행동들로 행동에 부여되는 의미가 커진다고 할 수 있다.

참고문헌 [4]도 이와 유사하게 분류하고 있으나 행동 단위에 있어 다소 다른 분류체계를 갖는다. 사용자의 행동을 Action, Activity, Behavior로 구분하고 있는데, Action이 가장 작은 단위의 행동이나 참고문헌 [3]에서 정의한 앉기, 서기, 걷기 등과 같은 기본 행동이라기보다 문 열기, 물체 옮기기, 조명

끄기, 걷기 등 보다 구체적인 행동을 타임 스탬프와 함께 Action으로 정의하였다. Activity는 ‘커피 마시기’, ‘회의 참석’, ‘화상 회의’, ‘이메일 보내기’ 등 목적을 갖고 이루어지는 일련의 Action 집합으로 정의된다. Behavior는 Action 또는 Activity의 시퀀스로 구성된다.

앞의 사례에서 보듯이 행동 인식을 위한 대상 분류와 정의는 각 연구의 목적과 특성에 맞추어 저수준의 행동으로부터 고수준의 행동으로 정의하여 사용하고 있다.

## 2. HAR 데이터셋 주요 특징

행동 인식(HAR: Human Activity Recognition) 기술 분야 연구를 위해서는 행동 데이터셋 구축이 선행되어야 한다. 그러나 행동 데이터셋 구축은 비용과 시간이 많이 들고, 기존 연구결과와 비교할 때 인식율의 객관성 확보가 쉽지 않기 때문에 많은 연구자는 기존에 발표된 공공 데이터셋을 이용하고 있다. 공공 데이터셋 중 스마트폰 또는 신체부착형(On-body) 센서를 사용하여 일상생활에서 데이터를 수집한 데이터셋으로는 UCSD ExtraSensory [5], UCI-HAR[6], mHealth[7]가 있다. 국내에서는 한국전자통신연구원(ETRI)이 이와 비슷한 방법으로 2018년부터 손목착용형 센서와 스마트폰

을 이용하여 업무, 학습, 회의, 통근/통학 등 일상 생활 중에 빈번히 발행하고 의미적으로 구분할 수 있는 17개 행동의 레이블 데이터를 수집하고, 행동이 시작하고 끝날 때까지의 시간, 자세, 장소, 휴대폰 위치, 같이 있는 사람 등 다양한 맥락 레이블을 같이 수집하여 데이터셋을 구축하였다[8]. 표 1은 이들 데이터셋을 피실험자 수, 수집 기간, 수집 데이터, 사용 디바이스 관점으로 비교한 표이다.

UCI-HAR은 스마트폰만을 데이터 수집 디바이스로 이용한 반면, mHealth는 스마트폰, 스마트워치 및 피부에 직접 부착하는 신체부착형 센서를 이용하였고, ExtraSensory와 ETRI 데이터셋은 스마트폰과 스마트워치를 함께 이용하였다.

UCI-HAR의 경우, 인식 대상이 되는 행동이 걷기, 앉기, 서기, 눕기, 계단 올라가기, 내려가기 등 기본 행동을 대상으로 하고 있다. ExtraSensory와 ETRI 데이터셋의 경우 기본 행동뿐만 아니라 라이프스타일을 알 수 있도록 실생활에서 자기보고 형식으로 고수준 행동(Behavior) 레이블을 함께 수집하였다.

## 3. 행동인식 기술

기존의 행동 인식을 위한 데이터셋은 대부분 제

표 1 행동 데이터셋 비교

데이터셋	피험자 수	수집 기간	수집 데이터	사용 디바이스
UCI-HAR[6]	30명	96분	Accelerometer, Magnetometer, Protocol Labels	Smartphone
mHealth[7]	10명	120분	Accelerometer, Gyroscope, Magnetometer, ECG, Video	On-body sensors (3IMUs, ECG)
ExtraSensory[5]	60명	7일	Accelerometer, Gyroscope, Magnetometer, GPS, Audio, Phone State, Self-report Labels	Smartphone, Smartwatch
ETRI dataset[8]	22명	28일	Accelerometer, Gyroscope, Magnetometer, GPS, Audio, Phone State, PPG, EDA, Skin Temp., Self-report Labels	Smartphone, Smartwatch

출처 Reproduced from [5-8].

한된 실험실 환경에서 지정된 행동을 수행하는 방식으로 진행되어왔다. 이러한 데이터를 기반으로 학습된 모델은 실생활에서 발생하는 다양한 타겟 행동 인식을 위한 적용에는 어려움이 있다. 사용자에 따른 행동 특징의 차이, 센서 디바이스에 따른 데이터 특징의 차이, 디바이스의 착용 위치 및 수집 환경의 다양성 등이 대표적인 예이다. 이러한 도메인 변화에 강건하게 동작하는 행동 인식 모델 개발을 위해 비지도학습 기반 도메인 적응(UDA: Unsupervised Domain Adaptation) 방법을 적용한 행동 인식 기술 관련 연구들이 진행되고 있다. 참고문헌 [9]는 스마트워치, 헬스케어 밴드, 이어폰, 스마트폰 등 디바이스들의 다양한 위치와 자유도를 착용 다양성에 의한 도메인 변화(Domain Shift) 문제로 해석하고, 이를 해결하기 위해 UDA 딥러닝 방법을 적용하였다. 소스 도메인과 타겟 도메인의 특징 공간이 서로 비슷하도록 조정하는 것을 목표로 하며, 데이터 증강(Data Augmentation) 기법, 특징 매칭(Feature Matching) 기법과 혼동 극대화(Confusion Maximization) 기법을 적용하여 결과를 비교하였다. 실험 결과, 특징 매칭 기법이 타겟 도메인 적응 이후에도 소스 도메인 학습 모델의 무결성을 저해하지 않으면서도 도메인 간의 관계를 학습하는 데 나은 성능을 보였으며, 타겟 도메인 분류 성능을 향상시켰다. 데이터 증강 기법은 디바이스를 신체에 착용하는 다양성에 적응하는데 효과적인 반면, 혼동 극대화 기법은 착용 위치 차이가 큰 경우에 효과적인 것으로 나타났다. 또한, 강인한 도메인 적응 모델을 위해서는 여러 착용 위치에서의 데이터와 자유도가 높은 센서 착용 위치에서의 데이터가 결정적인 것으로 확인하였다.

참고문헌 [10]은 기존 연구가 타겟 도메인 데이터셋이 미리 존재하는 비실시간의 도메인 적응 시

나리오를 가정한 것에 반하여, 실시간으로 타겟 도메인 데이터셋에 적용 가능한 실시간/온라인 도메인 적응 기법의 중요성을 강조하였다. 사용자별 데이터를 소스 도메인으로 보았을 때 행동 인식의 개인화에서 다중 소스 도메인 적응 문제로 적용하였으며, 도메인 조정을 위한 배치 정규화 레이어를 사용한 기법을 제안하였다. 이 기법은 실시간 분석을 진행하기 때문에 대용량의 데이터 저장을 필요로 하지 않는다. 또한, 일정한 컴퓨터 자원을 효율적으로 사용하기 때문에 잠재적으로 많은 양의 실시간 데이터를 처리할 수 있다는 장점이 있다. 그러나, 실시간 도메인 적응을 위한 최적의 하이퍼파라미터 선택을 위해서 그리드 탐색 기반 대규모 실험이 필요하다는 단점이 있다.

한편으로는, 레이블된 데이터셋의 한계를 극복하고 학습에 사용되는 데이터의 다양성과 복잡성을 향상시키기 위해서 레이블되지 않은 데이터도 함께 활용하는 준지도학습 방법을 적용한 연구들이 진행되어왔다. 이런 연구에서는 대용량의 도메인 데이터셋으로 네트워크 학습을 진행한 다음, 다른 태스크 혹은 도메인 데이터를 테스트 데이터로 활용하여 미세 조정을 수행하는 전이학습 방법이 널리 사용되었다. 참고문헌 [11]은 기존 행동 인식 연구에서 사용되지 않은 준지도학습 방법 중 하나인 자기학습(Self-training)을 도입하였다. 이 연구에서는 레이블된 데이터셋은 그 크기가 제한된 반면, 레이블되지 않은 데이터셋은 상대적으로 많이 존재하는 점에 착안하여 레이블되지 않은 데이터도 학습에 함께 활용하는 방법을 제안하였다. 이 방식은 지도학습과 준지도학습 방법의 추론 복잡도를 증가시키지 않으면서도 더 높은 성능을 보이는 것으로 확인되었지만, 레이블되지 않은 데이터셋을 학습에 활용하는 것일 뿐 해당 데이터의 레이블을 추론하는 것이 아니라는 한계가 있다. 참고문

현 [12]는 데이터의 레이블링을 위한 시간과 노력, 그리고 보안과 관련된 문제를 완화하기 위해 도메인 지식을 필요로 하지 않는 비지도학습 기반의 멀티 태스크 딥 클러스터링 행동 인식 방법을 제안하였다. 이 연구는 기존의 딥 클러스터링 방식이 CNN에 기반하였던 것에 비해 시계열 센서 데이터의 시공간적 연관성을 반영하는 CNN-LSTM 구조의 인코더를 도입하였다는 특징이 있다. 또한, 딥 클러스터링과 멀티 태스크 학습 기법을 통합하여 군집화에 적합하면서도 행동 분류를 목적으로 하는 데이터 표현형 생성이 가능하다는 장점이 있다. 하지만, 군집화 기반의 방식이기 때문에 데이터의 분포가 불균형을 이룰 경우 편향된 결과를 도출하게 되므로, 행동 클래스별 데이터의 수가 동일하게 설정되어야 하고 전체 데이터 중 사용자별 데이터 분포 또한 균등하게 설정하는 전처리가 필요하다.

최근 스마트 기기를 이용하여 개인의 일상을 모델링하고 행동패턴을 분석하는 연구 또한 활발하다[13]. 특히 개인의 채팅 기록, 접속 기록, 구매 기록 등 라이프로그 데이터를 지식 그래프, 시맨틱 네트워크 구조로 나타내어 개인의 일상을 모델링하고자 하는 시도가 있는데, 그래프 형식의 라이프로그 데이터는 노드와 엣지를 활용함으로써 노드 간의 데이터 관계를 나타내어 개인의 일상에 대한 의미정보를 표현하는 데 적합하다[14]. Google은 빅데이터의 형태로 저장된 수많은 정보를 표현하기 위해 노드, 엣지의 연결구조를 갖는 지식 그래프를 생성하였다. 구축된 지식 그래프를 기반으로 검색 기록 사이의 연관성 분석 등의 작업을 통하여 검색 성능이 향상되었다. 개인화된 대화 시스템으로 자연스러운 대화를 수행하는 구글 어시스턴트에서는 채팅 로그, 라이프로그 데이터를 시맨틱 그래프로 표현하여 학습하였다. 그래프 구조를 이용

한 행동패턴 데이터 분석 기술은 노드와 엣지로 연결된 저장 구조를 이용하여 연관 정보를 추출함으로써 패턴을 분석할 수 있다. 하지만 그래프의 가변적인 크기, 그래프마다 다른 근접 이웃 등 그래프 데이터의 복잡성으로 인해 전통적인 탐색 알고리즘을 사용하여 유의미한 특징을 뽑아내기에는 한계가 있다[15]. 딥러닝 그래프 임베딩 기술은 그래프를 저차원의 벡터로 표현하여 문제를 해결하지만 블랙박스 특성으로 인해 유의미한 패턴 분석에 어려움이 따르고 도메인 지식 활용에 한계가 존재한다. 따라서 그래프 신경망을 통한 데이터 기반 접근방법과 도메인 지식을 통한 심볼릭 기반 접근방법을 결합하는 뉴로-심볼릭 결합기술이 제안되고 있다.

#### 4. 라이프로그 이벤트 분할 기술

변화시점감지(CPD: Change Point Detection)는 시계열 데이터의 추세와 특성이 변화하는 시점을 감지함으로써 레이블되지 않은 데이터에서 의미를 추출하는 데 많이 사용되는 기법이다. 자동화 시스템에서 시스템 폴트로 인한 사람의 개입이 필요한 문제시점 탐지에서부터 행동 인식, 이동 경로 예측, 제스처 인식, 라이프로그 등의 도메인에서 시계열 데이터의 시간에 따른 분할 이벤트 감지, 이상징후 탐지 등 시스템의 동작 특성을 이해하는 데 활용되고 있다. 기존 연구에서 변화 시점은 시계열 데이터의 다양한 특성, 그중에서도 시간에 따른 시계열 형태 및 연속성 통계적 분포(Auto Regressive Model, Kernel Function 등) 변화 등에 기반하여 결정하였다. 그러나 어떤 응용 도메인에 적용되는지에 따라 데이터의 의미 구간을 결정하는 방법이 달라지기 때문에 그 방법을 일반화하는 데 한계가 있다. 이 절에서는 라이프로그 데이터에 적용 가능한

다양한 이벤트 분할 기법과 관련된 선행연구에 대해 분석한다.

참고문헌 [16]은 변화시점 감지 문제에 자기지도 학습 기법 기반 대조예측코딩(CPC: Contrastive Predictive Coding)을 적용한 시계열 변화점 감지 방법을 제안하였다. 이는 시계열 데이터의 지역적 상관관계를 활용하여 연속적인 시계열 데이터 간의 공통적인 데이터 표현형을 최대화하면서, 시간적으로 떨어져 있는 상관관계가 낮은 시계열 데이터 간에는 공통적인 데이터 표현형을 최소화하는 자기지도 방식으로 학습을 진행하였다. 이 연구에서는 성능 평가를 위해 USC-HAD 데이터셋[17]에서 30개의 행동에 해당하는 데이터를 6명의 사용자로부터 임의로 추출하여 다시 임의로 엮는 방식으로 총 30개의 변화시점을 생성하여 실험에 사용하였다. 하지만 이렇게 인위적으로 조합한 데이터의 변화시점은 연속적이지 않으므로 실제 데이터에서 나타나는 연속적인 변화시점과는 차이가 있다.

참고문헌 [18]은 많은 양의 레이블이 없는 다양한 형태의 연속적인 시계열 데이터셋을 사람의 개입 없이 자동으로 이산적 세그먼트의 시퀀스로 분할하는 시계열 분할 기법을 제안하였다. 해당 기법을 적용하면 연속적인 시계열 데이터를 비지도 학습 기반으로 작은 단위의 세그먼트로 분할하는 전처리가 가능하다. 그 결과 데이터에 대한 이해도를 높여 데이터 특징을 추출하는 데 도움을 줄 수 있다. 기존 통계 기반 데이터 분할 연구는 시계열 데이터 자체에 많은 제약 사항을 가정하는데, 이 연구에서는 이러한 한계를 극복하기 위해 시계열 데이터의 엔트로피와 시간적 특성을 동시에 반영하는 하이브리드 방식의 변화시점 감지 알고리즘을 제안하였다.

### III. 감정 인식 기술 동향

#### 1. HCI 감정 인식 기술

인간의 감정은 개인적 경험, 생리적 또는 행동적 반응, 또는 대화 시의 반응과 같은 여러 요소들로 구성되어 표현되는 감정 상태를 일컫는다[19]. 인간은 일상생활에서 상대의 감정을 파악하기 위해 상대의 음성, 표정, 행동 정보 중 하나 이상의 두드러진 특징과 여러 문맥적 정보와 혼합하여 파악하게 된다. 기존 연구에서는 개인의 감정 상태를 시간적 구간 특성에 관련하여 감정을 분류하고 정의하였다[20]. HCI 분야에서 인식하는 사람의 감정은 대체적으로 분(Minutes) 또는 시간(Hours) 구간으로 짧은 시간에 강하게 표현되는 ‘full-blown’ 감정을 지칭하는 경우가 많다. ‘Mood’ 또는 ‘Emotional Disorders’는 상대적으로 오랜 시간 또는 수개월 동안 지속되는 심리적 상태를 지칭한다.

HCI 분야에서 기계가 상대의 감정을 파악할 때도 일반적인 감정인식에서 사용하는 정보에 기반해야 한다. 기계는 인간과의 상호작용 중에 발생하는 사용자의 직접적 반응 내용(What)이 전달되는 명확한(Explicit) 채널, 또는 사용자의 반응 방식(How)이 전달되는 내포적인(Implicit) 채널로부터의 정보를 통해 상대의 감정을 인식한다. 음성 정보에 기반하는 감정인식 기술을 가정할 경우, 명확한 감정 채널의 정보로는 화자가 말하는 내용이 될 수 있고, 내포적인 채널로 전달되는 정보는 화자의 말하는 방식이 해당된다.

HCI 분야에서의 감정인식은 특정 감정 분류로 구분되는 이산적(Discrete) 감정 모델[21] 또는 긍부정도(Valence), 각성도(Arousal), 파워(Power) 등의 감정 속성을 소수 단위의 레벨로 정의하는 차원적(Dimensional) 감정 모델[22]에 기반한다. 이산적 감

정 모델 기반 감정인식에서는 슬픔, 기쁨, 공포, 분노, 혐오, 그리고 놀람과 같은 감정 레이블로 인식한다.

## 2. 음성 감정 인식 기술

음성신호에 기반하는 감정 인식(SER: Speech Emotion Recognition)은 사용자가 발화한 음성신호를 처리하여 이산 감정 모델 또는 차원적 감정 모델을 이용하여 감정 레이블을 인식한다. 최근에는 음성신호의 유효한 특징을 추출하고, 추출된 특징을 심층 기계학습 모델을 통해 학습하는 감정인식 연구가 활발하다.

표 2는 기계학습을 위해 음성신호로부터 추출되는 음성특징의 분류를 보인다. 운율(Prosodic) 분류에서는 음성의 억양(Intonation), 리듬(Rhythm)과 관련된 음성특징을 포함한다. 스펙트럼(Spectrum) 관련 특징은 시간 도메인의 음성신호를 주파수 도메인의 특징으로 변환한 MFCC(Mel Frequency Cepstral Coefficients), LPCC(Linear Prediction Cepstral Coefficients) 등을 포함한다. 음성 품질에 관련된 특징에는 jitter, shimmer, HNR과 같은 음성의 물리적 특징을 포함한다. TEO(Teager Energy Operator) 특징은 스트레스와 같은 요인에 의해 발화 시 성대의 근육에 가해지는 에너지를 정의하는 물리적인 음성특징에 관한 것이다[23].

기계학습 기반 음성 감정인식 모델을 구축하기

표 2 음성 특징 분류

Prosodic	Spectral	Voice Quality	TEO Features
Pitch Energy Duration	MFCC, LPCC LFPC, GFCC Formants	Jitter Shimmer HNR, Amplitude	TEO-FM-Var TEO-Auto-Env TEO-CB-Auto-Env

출처 Reproduced from [19].

위해서는 인식하고자 하는 감정 모델의 레이블이 부여된 음성 감정 데이터셋이 필요하다. 음성 감정 데이터셋의 구축은 감정 레이블의 분류와 강도를 평가함에 있어 개인이 갖는 주관성, 음성 감정 신호에 대한 개인적, 문화적 또는 환경적 다양성으로 인해 음성 감정 데이터 수집에 많은 시간과 비용이 필요하다.

현재 공개되어 있는 감정 데이터셋은 그 수집 방법에 따라 특정 감정 상황에 대해 연기하거나 (Acted, Simulated), 오디오 또는 비디오 자극을 통해 사용자로부터 특정 감정을 유도하는 방법 또는 자유발화(Natural Conversation) 수집 방법으로 구분된다. 대표적으로 IEMOCAP[24]과 KESDy18[25] 데이터셋은 연기된 데이터셋이며, eNTERFACE'05 audio-visual[26]은 유도의 방식으로 수집된 데이터셋, SAMAIN[27]은 자유발화 데이터셋으로 구분된다. 수집된 데이터셋에 대한 감정 레이블의 태깅은 수집 대상자, 즉 화자의 자기보고 방식 또는 외부 감정평가자의 평가에 의한 감정 레이블 부여의 방식에 따라 부여된다.

최신의 음성 감정인식 결과는 대부분은 RNN, CNN에 기반하는 심층학습 기반의 학습모델에 기반하여, 단일 네트워크 또는 복수의 네트워크를 혼합한 앙상블 구조로 구현한다. 최근에는 특정 음성 감정 소스 데이터셋을 기반으로 지도(Supervised) 또는 비지도(Unsupervised) 방식으로 학습된 훈련모델을 활용하여 추출한 특징 정보를 새로운 타겟 도메인에서의 감정 학습모델에 이용하는 전이학습 기반의 음성 감정 모델에 대한 연구가 활발히 수행되고 있다[28-30].

## 3. 무드 인식 기술

웨어러블 디바이스로 수집 가능한 많은 라이프

로그 데이터를 바탕으로 개인의 감성 또는 무드(Mood)를 인식하거나 예측하는 기술들도 연구가 활발히 진행되고 있다.

참고문헌 [31]은 스마트폰과 스마트워치를 이용하여 HR(Heart Rate), GSR(Galvanic Skin Response) 데이터를 수집하고, PSS(Percieved Stress Scale)를 5점 척도로 하루 두 번 응답받아 부정적 감성(스트레스)을 감지하는 기술을 구현하였다. 사용자가 자각하지 않은 상태에서 생리반응 신호를 수집하여 스트레스를 모니터링하고 피드백함으로써 질병 예방과 같은 효과를 목적으로 하였다. 참고문헌 [32]는 사용자를 심리적으로 이해하기 위해 라이프로그 데이터를 사용하였다. 아침 기상 직후에 무드(Happy, Peaceful, Depressed, Anxious)를 수집하여 이를 긍부정도(Arousal/Valence)로 분류하고 신체활동과 바이오메트릭스를 함께 이용하여 무드를 예측하는 방법을 제시하였다. 참고문헌 [33]은 회사에 근무하는 60명의 사무실 노동자들로부터 주중 수면 상태에 대한 웨어러블 센서 데이터와 건강지수를 측정하기 위한 기상 직후 DAMS(Depression and Anxiety Mood Scale) 질문지 응답을 수집하고, 이를 바탕으로 우울, 불안 무드를 인식하는 모델을 제시하였다.

이러한 연구들은 무드 예측을 위해 Fibit, MS Band2, Affective Q Sensor, E4 등과 같은 손목 착용형 디바이스를 사용하였다. 착용 디바이스로부터 HR, GSR, SKT(Skin Temperature), 가속도 등의 데이터를 수집하였다. 라이프로그 레이블로는 수면 정보를 비롯하여 행동이나 상황정보와 관련된 특징들을 수집하여 사용하고 있다. 특히, 수면은 개인의 감성 또는 무드에 큰 영향을 미치는 것으로 보고되고 있어 많은 연구에서 수면과 관련된 특징들을 추출하여 활용하고 있다.

개인의 감성 또는 무드는 응용에 따라 목표로 하

는 무드가 다양하게 존재하고 있다. 대표적으로 신체적·정신적 건강에 가장 영향을 많이 미치는 스트레스를 중요한 개인의 감성으로 다루고 있었고, 이외에도 다음 날 아침 기상 직후의 기분을 예측하려는 연구들도 있다. 이 분야 연구 역시 많은 사람을 대상으로 양질의 데이터를 대량으로 확보하는 것이 인식의 정확도를 높이는 문제와 직결되는 만큼 대용량, 고품질의 라이프로그 데이터를 확보하는 기술에 대한 연구가 시급한 연구과제로 보고되고 있다.

#### IV. 활용 서비스 분야 동향

최근 의료계에서는 스마트폰을 사용하는 패턴을 분석하여 그 사람의 건강상태나 질병 유무를 파악하는 기술에 주목하고 있다. 이러한 스마트폰 기반 디지털 표현형(Digital Phenotype)은 GPS 데이터, 가속도계 데이터, 통화 기록 및 메시지 로그, Wi-Fi 및 블루투스 연결 기록과 스마트폰 마이크로 수집한 음성 샘플 등을 통해 얻을 수 있어 모바일 데이터를 이용한 사용자 행동 연구에 활용하고 있다 [34].

의료 서비스 분야에서는 ‘사용자 데이터’를 사용자가 의식적으로 기록하는 능동적(Active) 데이터와 사용자가 의식하지 않을 때 디바이스가 스스로 데이터를 수집하여 기록하는 수동적(Passive) 데이터로 나누어 정의한다. 수동적 데이터를 통해서는 기억 과정의 편향이나 인지적 추론의 편향 등 자기 보고 방식의 문제점[35]을 벗어나 최대한 객관적인 사용자 정보를 얻을 수 있다. 모바일 데이터에서 추출한 메타데이터를 통해서는 사용자의 건강과 병증에 대한 주요 정보를 얻을 수 있다. 예를 들어, GPS 데이터에서 추출한 머문 지점 및 움직임 메타데이터는 우울증 환자가 집과 직장과 같은 별



개의 위치에서 시간을 얼마나 보내고, 얼마나 다양한 곳을 방문하는지와 같은 정보를 나타내는 데 사용된다. 통화 기록은 개인의 사회성을 나타내는 정보로 활용될 수 있다. 소셜 마커는 양극성 장애 환자의 우울증과 조증을 오가는 환자의 양상을 나타낼 수 있다. 스마트폰으로부터 얻은 음성 샘플은 기분을 감지하는 데 사용되거나 언어 장애를 특징으로 갖는 신경 장애에 대한 예후를 진단하는 데 사용될 수 있다. 마지막으로, 가속도계 데이터는 다양한 환자의 신체 이동 패턴과 대사 지출을 정량화하는 데 광범위하게 사용될 수 있다. 가속도계 데이터에 도메인 지식을 반영한 메타데이터를 이용한 사례로는, 비자발적 근육 수축으로 고통받는 환자의 떨림을 정량화한 경우이다. 스마트폰 센서와 통화 패턴 등의 데이터는 사용자의 다양한 사회적·행동적 정보 제공의 잠재력을 가지고 있으나, 데이터의 다양성과 복잡성을 충분히 고려하지 않으면 원하는 목적을 달성하기 어렵다. 따라서 이와 같은 정보를 얻기 위해서는 원시데이터를 사용 목적에 맞게 전처리하여 메타데이터를 추출하는 과정이 필요하다.

디지털 치료제로 가장 대표적인 기업은 Pear Therapeutics[36]이다. Pear Therapeutics는 불면증을 치료하기 위한 디지털 치료제인 Somryst를 개발하고 미국 FDA 승인을 획득하였다. Somryst는 스마트폰 앱을 이용한 인지행동치료로 수동적 데이터의 사용보다는 사용자가 직접 본인의 상태를 입력하는 능동적 데이터를 주로 활용하고 있다. 이를 이용하여 실제로 불면증을 개선하는 효과를 거두었다는 임상실험 결과도 발표되었다[37]. 이와 비슷하게 사용자가 직접 기록에 참여하는 능동적 데이터를 이용하는 서비스로는 BlueSignum[38]의 하루콩을 예로 들 수 있다.

수동적 데이터를 활용하는 가장 선도적인 기

업은 Google의 자회사 Verily 소속의 신경과학자 Thomas Insel이 공동창업한 Mindstrong Health[39]라고 할 수 있다. Mindstrong Health는 스마트폰 사용 패턴을 이용해 우울증의 재발, 자살 같은 징후를 포착하거나 예측하는 서비스를 제공한다. 스마트폰을 타이핑하는 방식, 스크롤을 내리는 방식, 화면을 터치하는 방식에 대한 데이터를 수집하여 기계학습을 활용하는 방식이다[40].

디지털 치료제와 관련한 산업은 이제 막 태동한 산업으로 앞으로 다양한 디지털 데이터를 이용한 서비스들이 등장할 것으로 예상된다.

## V. 결론

본고에서는 실험실 상황이 아닌 실생활에서 사용자 행동을 인식하기 위한 데이터 수집 관점에서 현재 많은 연구에 인용되고 있는 공공 데이터셋의 주요 특징과 한계를 비교하여 알아보았다. 이를 통해 기존의 많은 연구가 실생활에서 사용자 행동인식을 위한 데이터 수집에 어려움을 겪고 있는 것을 확인할 수 있었다. 또한, 실생활 적용을 위한 행동 인식 기술의 최근 동향과 시계열 데이터의 이벤트 분할 방법에 대한 연구 동향, 음성 특징 기반의 감정 인식 기술과 라이프로그 데이터로부터 무드를 예측하는 연구 동향에 대해 살피고 주요 이슈를 알아보았다.

연구 동향 분석 결과 실생활 적용을 위해 대량의 고품질 데이터를 확보하기 위해서는 항목 선정과 구성을 고려한 레이블 설계가 중요하다는 것을 알 수 있었다. 또한, 데이터 수집에 많은 어려움이 있는 상황에서 부족한 데이터를 활용하여 행동과 감정 인식 알고리즘을 개선하는 방향의 노력이 지속적으로 이루어지고 있음을 확인하였다.

휴먼이해 인지컴퓨팅 기술은 향후 인간과 교감

이 가능한 인공지능의 핵심 기술로 적용되어 신체적·정신적 건강을 지켜 삶의 질을 높이는 서비스를 실현해 주는 기술로 활용될 가능성이 크므로, 이와 관련된 연구에 집중적인 노력과 지원이 필요한 시점이라고 판단된다.

## 약어 정리

ADL	Activities of Daily Living
CNN	Convolutional Neural Network
CPD	Change Point Detection
HAR	Human Activity Recognition
HCI	Human Computer Interface
HNR	Harmonics to Noise Ratio
IMU	Inertial Measurement Unit
RNN	Recurrent Neural Network
UDA	Unsupervised Domain Adaptation

## 참고문헌

- [1] 캐서린 러브데이, "나는 뇌입니다," 행성B, 2016, pp. 1-332.
- [2] 리사 펠드먼 배럿, "감정은 어떻게 만들어지는가?," 생각연구소, 2017, pp. 219-244.
- [3] A.A. Chaaoui, P. Climent-Pérez, and F. Flórez-Revueita, "A review on vision techniques applied to human behaviour analysis for ambient-assisted living," *Expert Syst. Appl.*, vol. 39, no. 12, 2012, pp. 10873-10888.
- [4] N.D. Rodriguez et al., "A fuzzy ontology for semantic modelling and recognition of human behaviour," *Know.-Based Syst.*, vol. 66, 2014, pp. 46-60.
- [5] Y. Vaizman et al., "ExtraSensory app: Data collection in-the-wild with rich user interface to self-report behavior," in *Proc. HI Conf. Hum. Factors Comput. Syst.*, (Montréal, Canada), Apr. 2018, pp. 1-12.
- [6] D. Garcia-Gonzalez et al., "A public domain dataset for human activity recognition using smartphones," in *Proc. Eur. Symp. Artif. Neural Netw., Comput. Intell. Mach. Learn. (ESANN)*, (Bruges, Belgium), Apr. 2013, pp. 437-442.
- [7] A. Prasad et al., "Provenance framework for mHealth," in *Proc. Int. Conf. Commun. Syst. Netw. (COMSNETS)*, (Bangalore, India), Jan. 2013.
- [8] S. Chung et al., "Real-world multimodal lifelog dataset for human behavior study," *ETRI J.*, Dec. 8, 2021.
- [9] Y. Chang et al., "A systematic study of unsupervised domain adaptation for robust human-activity recognition," *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.*, vol. 4, no. 1, 2020, pp. 1-30.
- [10] A. Mazankiewicz, K. Böhm, and M. Berges, "Incremental real-time personalization in human activity recognition using domain adaptive batch normalization," *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.*, vol. 4, no. 4, 2020, pp. 1-20.
- [11] C.I. Tang et al., "SelfHAR: Improving human activity recognition through self-training with unlabeled data," *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.*, vol. 5, no. 1, 2021, pp. 1-30.
- [12] H. Ma et al., "Unsupervised human activity representation learning with multi-task deep clustering," *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.*, vol. 5, no. 1, 2021, pp. 1-25.
- [13] B. Elser and A. Montresor, "An evaluation study of BigData frameworks for graph processing," in *Proc. IEEE Int. Conf. Big Data*, (Silicon Valley, CA, USA), Oct. 2013, pp. 60-67.
- [14] X. Zhan and S.V. Ukkusuri, "A graph-based approach to measuring the efficiency of an urban taxi service system," *IEEE Trans. Intell. Transportation Syst.*, vol. 17, no. 9, 2016, pp. 2479-2489.
- [15] Z. Wu et al., "A comprehensive survey on graph neural networks," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 32, no. 1, 2020, pp. 1-21.
- [16] S. Deldari et al., "Time series change point detection (CPD) with self-supervised contrastive predictive coding (CPC)," in *Proc. Web Conf.*, (Ljubljana, Slovenia), Apr. 2021, pp. 3124-3135.
- [17] M. Zhang and A.A. Sawchuk, USC-HAD: A daily activity dataset for ubiquitous activity recognition using wearable sensors," in *Proc. ACM Conf. Ubiquitous Comput.*, (Pittsburgh, Pennsylvania), 2012, pp. 1036-1043.
- [18] S. Deldari et al., "ESPRESSO: Entropy and ShaPe aware timE-Series SegmentatiOn for processing heterogeneous sensor data," *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.*, vol. 4, no. 3, 2020, pp. 1-24.
- [19] M.B. Akçay and K. Oğuz, "Speech emotion recognition: Emotional models, databases, features, preprocessing methods, supporting modalities, and classifiers," *Speech Commun.*, vol. 116, 2020, pp. 56-76.
- [20] R. Cowie et al., "Emotion recognition in human-computer interaction," *IEEE Signal Process. Mag.*, vol. 18, no. 1, 2001, pp. 32-80.
- [21] P. Ekman and D. Keltner, "Universal facial expressions of emotion," *Seegerstrale U P Molnar P Eds Nonverbal Commun. Nat. Meets Cult.*, vol. 27, 1997, p. 46.
- [22] J.A. Russell and A. Mehrabian, "Evidence for a three-factor theory of emotions," *J. Res. Personal.*, vol. 11, no. 3, 1977, pp. 273-294.
- [23] H.M. Teager and S.M. Teager, "Evidence for nonlinear sound

- production mechanisms in the vocal tract,” in *Speech Production and Speech Modelling*, Springer, Dordrecht, Netherlands, 1990, pp. 241-261.
- [24] C. Busso et al., “IEMOCAP: Interactive emotional dyadic motion capture database,” *Lang. Resour. Eval.*, vol. 42, no. 4, 2008, pp. 335-359.
- [25] K.J. Noh et al., “Multi-path and group-loss-based network for speech emotion recognition in multi-domain datasets,” *Sensors*, vol. 21, no. 5, 2021, p. 1579.
- [26] O. Martin et al., “The eINTERFACE’05 audio-visual emotion database,” in *Proc. Int. Conf. Data Eng. Workshops (ICDEW’06)*, (Atlanta, GA, USA), Apr. 2006, pp. 1-8.
- [27] G. McKeown et al., “The semaine database: Annotated multimodal records of emotionally colored conversations between a person and a limited agent,” *IEEE Trans. Affect. Comput.*, vol. 3, no. 1, 2011, pp. 5-17.
- [28] M.N. Stolar et al., “Real time speech emotion recognition using RGB image classification and transfer learning,” in *Proc. Int. Conf. Signal Process. Commun. Syst. (ICSPCS)*, (Surfers Paradise, Australia), Dec. 2017, pp. 1-8.
- [29] W. Dai et al., “Modality-transferable emotion embeddings for low-resource multimodal emotion recognition,” in *Proc. ACL 21*, (Suzhou, China), Dec. 2020, pp. 269-280.
- [30] S. Latif et al., “Transfer learning for improving speech emotion classification accuracy,” in *Proc. Interspeech*, (Hyderabad, India), Sept. 2018, pp. 257-261.
- [31] N. Jaques et al., “Predicting tomorrow’s mood, health, and stress level using personalized multitask learning and domain adaptation,” *Proc. Mach. Learn. Res.*, vol. 48, 2017, pp. 17-33.
- [32] C. Dobbins et al., “A lifelogging platform towards detecting negative emotions in everyday life using wearable devices,” in *Proc. IEEE Int. Conf. Pervasive Comput. Commun. Workshops (PerCom Workshops)*, (Athens, Greece), Mar. 2018, pp. 306-311.
- [33] P. Soleimaninejad et al., “THIR2 at the NTCIR-13 lifelog-2 task: Bridging technology and psychology through the lifelog personality, mood and sleep quality,” in *Proc. NTCIR Conf. Eval. Inf. Access Technol.*, (Tokyo, Japan), Dec. 2017.
- [34] J.P. Onnela, “Opportunities and challenges in the collection and analysis of digital phenotyping data,” *Neuropsychopharmacology*, vol. 46, no. 1, 2021, pp. 45-54.
- [35] M. Allemand and M.R. Mehl, “Personality assessment in daily life: A roadmap for future personality development research,” in *Personality Development Across the Lifespan*, Academic Press, London, UK, 2017, pp. 437-454.
- [36] <https://peartherapeutics.com/>
- [37] C.M. Morin, “Profile of somryst prescription digital therapeutic for chronic insomnia: Overview of safety and efficacy,” *Expert Rev. Med. Devices*, vol. 17, no. 12, 2020, pp. 1239-1248.
- [38] <https://www.bluesignum.com/>
- [39] <https://mindstrong.com/>
- [40] P. Dagum, “Digital biomarkers of cognitive function,” *NPJ digital medicine*, vol. 1, no. 1, 2018, pp. 1-3.