

스마트폰 기반 행동인식 기술 동향

Trends in Activity Recognition Using Smartphone Sensors

김무섭 (M.S. Kim, gomskim@etri.re.kr)	웨어러블컴퓨팅연구실 책임연구원
정치윤 (C.Y. Jeong, iamready@etri.re.kr)	웨어러블컴퓨팅연구실 책임연구원
손종무 (J.M. Sohn, jmsohn@etri.re.kr)	웨어러블컴퓨팅연구실 책임연구원
임지연 (J.Y. Lim, kusses@etri.re.kr)	웨어러블컴퓨팅연구실 선임연구원
정승은 (S.E. Chung, schung@etri.re.kr)	웨어러블컴퓨팅연구실 연구원
정현태 (H.T. Jeong, htjeong@etri.re.kr)	웨어러블컴퓨팅연구실 책임연구원
신형철 (H.C. Shin, shin@etri.re.kr)	웨어러블컴퓨팅연구실 책임연구원/실장

Human activity recognition (HAR) is a technology that aims to offer an automatic recognition of what a person is doing with respect to their body motion and gestures. HAR is essential in many applications such as human-computer interaction, health care, rehabilitation engineering, video surveillance, and artificial intelligence. Smartphones are becoming the most popular platform for activity recognition owing to their convenience, portability, and ease of use. The noticeable change in smartphone-based activity recognition is the adoption of a deep learning algorithm leading to successful learning outcomes. In this article, we analyze the technology trend of activity recognition using smartphone sensors, challenging issues for future development, and a strategy change in terms of the generation of a activity recognition dataset.

* DOI: 10.22648/ETRI.2018.J.330310

* 본 연구는 2018년도 정부(과학기술정보통신부)의 재원으로 정보통신기술진흥센터의 지원을받아 수행된 연구임[1711055381, 신체기능의 이상이나 저하를 극복하기 위한 휴먼 청각 및 근력 증강 원천기술 개발].



본 저작물은 공공누리 제4유형
출처표시+상업적이용금지+변경금지 조건에 따라 이용할 수 있습니다.

- I. 서론
- II. 스마트폰 기반
 행동인식 기술 동향
- III. 행동인식을 위한
 데이터베이스 동향
- IV. 결론

I. 서론

사용자 행동 인식(HAR: Human Activity Recognition) 기술은 다양한 센서를 활용하여 사람의 모션이나 제스처와 관련된 정보를 수집하고 해석하여 행동을 인식하는 기술을 의미한다. 사용자 행동은 다른 사람과 효과적인 소통을 위한 중요한 정보가 되며, 특히 HCI(Human-Computer Interaction), HRI(Human-Robot Interaction) 등 사람과 기술이 소통하기 위한 필수적인 정보를 제공해준다. 또한, 사람의 행동을 통하여 그 사람의 건강 상태를 유추할 수 있는 많은 정보를 제공하기 때문에 헬스케어 분야에서 많이 활용되고 있으며, 영상 감시시스템에서는 다양한 위협을 탐지하기 위한 목적으로 행동인식 정보를 사용하고 있다. 따라서 다양한 환경과 목적으로 사용자 행동 인식에 관한 연구가 진행되고 있으며, 기술 분석 기관인 Gartner는 행동 인식 기술을 차세대 스마트폰 시장을 견인할 수 있는 핵심 기술 중 하나로 전망하고 있다[1].

사용자 행동 인식 기술은 분석하는 데이터의 종류에 따라 크게 영상분석 기반의 행동인식 기술과 온바디 센서 기반의 행동 인식 기술로 구분할 수 있다[2]. 영상분석 기반의 행동인식 기술은 RGB 카메라 또는 Depth 카메라로부터 수집되는 영상을 분석하여 사람의 행동을 인식하는 기술이며, 온바디 센서 기반의 행동인식 기술은 사용자에게 장착된 가속도 센서(Accelerometer), 자이로스코프(Gyroscope), 블루투스, 사운드 센서 등의 다양한 센서 정보를 분석하여 행동을 인식하는 기술이다. 최근 스마트폰은 광범위한 사용자를 확보하고 있으며, 관성측정장치를 비롯한 다양한 센서를 기본적으로 내장하고 있어 사용자에게 별도의 센서를 부착하지 않아도 행동 인식에 필요한 데이터를 수집 할 수 있는 장점이 있다. 따라서 스마트폰을 이용하여 사용자의 행동을 인식하는 연구가 활발하게 진행되고 있다.

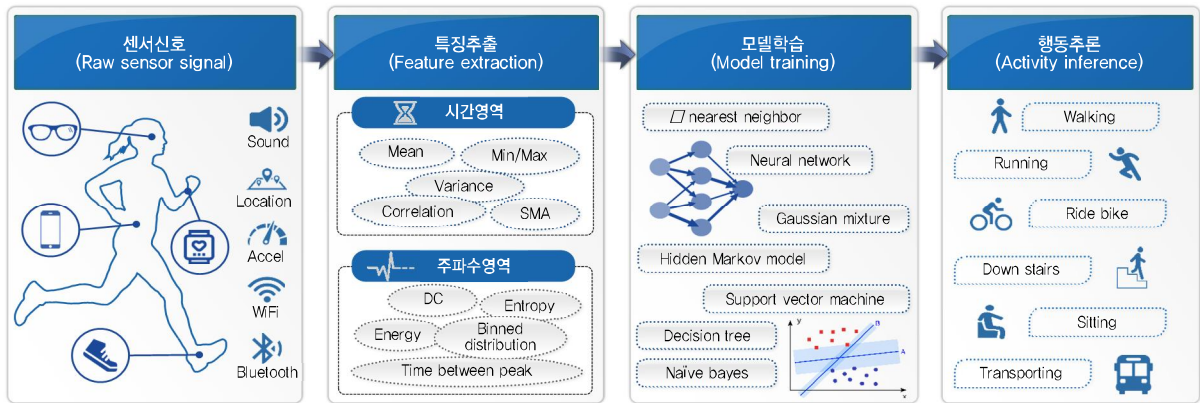
스마트폰을 이용한 행동인식 기술은 연구자가 직접

설계한 특징(Hand-Crafted Feature)에 기계학습 방법을 사용하여 행동을 인식하던 방법이 주로 사용되었으나, 최근에는 컴퓨터 비전, 음성인식 등에서 뛰어난 성능을 보여주는 딥러닝을 적용한 방법들이 많이 제안되고 있다. 따라서 본고에서는 온바디 센서 기반 행동인식 기술 중 스마트폰을 이용한 행동인식 기술의 동향과 이슈에 초점을 두고 기술 할 것이다. 또한, 행동인식 기술의 개발 및 성능 평가를 위해서는 필수적인 요소인 행동인식 데이터베이스 구축과 관련된 기술 동향을 추가로 분석하였다.

II. 스마트폰 기반 행동인식 기술 동향

스마트폰 기반 행동인식 기술은 스마트폰 또는 스마트폰과 연결된 다수의 센서에서 수집되는 신체 동작이나 기능 데이터를 활용하여 사람이 하는 일을 자동으로 인식하는 기술이다. 기존의 기술들은 행동인식을 위하여 신체에 웨어러블 센서나 RFID(Radio Frequency Identification) 태그 등의 장치들을 부착하여 데이터를 수집·활용 하였으나, 최근에는 스마트폰의 센서 데이터를 활용한 행동인식 기술들이 많이 개발되고 있다.

최근의 이러한 변화는 스마트폰 기술의 비약적인 발전과 함께 일상생활에서 스마트폰의 활용도가 높아진 점에서 원인을 찾을 수 있다. 먼저 디바이스 기술의 발전에 따라 개별적인 센서들의 정확도가 개선되고 소형화되면서 기존에 행동인식에 사용되던 다양한 센서들 상당수가 스마트폰에 기본적으로 내장되어 제공되고 있다. 따라서 행동인식을 위해 사람이 착용-부착해야 했던 다양한 장치들을 스마트폰으로 대체함으로써 사용자의 불편함을 감소시킬 수 있다. 또한, 스마트폰은 사용자에게 스마트폰 서비스를 원활하게 제공하면서도, 행동인식에 필요한 데이터는 사용자가 인지하지 못한 상태에서 자연스럽게 수집할 수 있기 때문에 사용자의 일상생활을 방해하거나 불편함을 초래하지 않는다.



(그림 1) 센서신호의 특징기반 행동인식 기술

기존 방법들의 경우, 다수의 센서를 사용하여 인식 알고리즘의 성능을 향상시킬 수 있지만, 사용자들이 느끼는 사용의 어려움과 착용의 불편함으로 일상생활에 적용하기에는 어려움이 있었다. 이러한 문제를 해결하기 위해 최근에는 스마트폰을 활용한 다양한 형태의 행동인식 기술이 개발되고 있다. 따라서 본 소절에서는 이러한 상황을 반영하여 스마트폰을 활용한 사용자 행동인식 기술의 동향에 대하여 분석하고자 한다.

일반적으로 행동인식을 위해서는 (그림 1)에서와 같이 센서 디바이스에서 신호의 수집, 수집된 신호에 대한 전처리 및 특징 추출, 분류기 모델의 학습, 그리고 분류기를 사용한 행동 추론의 과정이 필요하다.

스마트폰에서 제공하는 여러 센서 신호 중에서 행동인식에 가장 기본적이며 많이 활용되는 신호는 가속도 신호이다. 기존 연구[3]에 따르면 일반적으로 행동인식에 사용되는 특징들에 대하여 상관특징선택(CFS: Correlation Feature Selection) 알고리즘을 적용한 결과 60%의 특징들이 가속도 신호의 시간영역 특징들에 남아있다고 밝혀졌다. 또한, 자이로스코프(Gyroscope)와 자기력계(Magnetometer)에서 추출한 특징들을 추가로 사용하는 경우 대략 5% 정도의 성능 개선을 나타내는 것으로 알려져 있다. 즉, 행동인식에 활용되는 정보들의 상당 부분의 특징 정보가 가속도 신호에 집중되어 있으므로 가속도 신호가 많이 활용되고 있다.

일반적으로 가속도 센서만 사용하는 경우, 데이터가 측정된 위치(Location)나 방향(Orientation)에 따라 수집되는 데이터 값이 변화한다. 따라서 이러한 문제점을 해결하기 위한 다양한 방법들이 제안되었다. 가장 간단한 방법은 신체의 여러 곳에 다수의 독립적인 웨어러블 센서들을 부착하고 다수의 데이터를 수집하여 활용하는 것이다. [4]에서는 최대 5개의 2축 가속도 센서를 발목, 손목, 엉덩이, 팔, 허벅지의 위치에 부착하여 사용하였으나, 엉덩이와 손목 또는 허벅지와 손목의 위치에만 사용하는 경우보다 행동인식의 정확도가 크게 개선되지 않은 것으로 나타났다. 따라서 데이터의 수집을 위해 다수의 센서를 활용하기보다 데이터를 수집하는 위치의 효율적인 조합이 행동인식에 더 적절하다고 판단된다. 또한, 가속도 신호에 추가로 자이로스코프나 자기력계 등의 센서에서 수집한 데이터로부터 다양한 특징들을 추출하여 행동인식의 학습에 사용한다. 최근에는 기존의 걷기, 앉기 등의 단순하고 반복적인 행동 중심의 인식을 목표로 하는 저수준 행동인식에서 보다 정확하고 효율적인 행동인식을 위해 컨텍스트 정보(소리나 위치 등)를 포함하여 사용자의 행동을 인식하는 연구들이 발표되고 있다[5].

1. 데이터 수집 및 전처리

행동인식을 위해서는 가장 먼저 분석할 데이터의 수

집(Sampling)이 필요하다. 수집하는 데이터의 적절한 샘플링 주파수의 선택은 원하는 정확도, 목표 시스템의 특성 및 분석에 사용되는 데이터의 특징과 같은 요인들에 의해 결정되므로 적절한 샘플링 속도를 활용한 데이터 수집은 행동인식에서 중요하다. 행동을 인식하기 위하여 주파수 영역의 특징만 사용하는 경우를 예로 들자면, 샘플링 속도는 인간의 움직임에 관련된 모든 주파수를 캡처할 수 있을 정도로 충분히 높아야 한다. 선행 연구 결과들은, 측정된 모든 신체의 움직임은 20Hz 미만의 주파수 성분 내에 포함되어 있으며, 고속 푸리에변환(FFT: Fast Fourier Transform) 진폭의 98%가 10Hz 이하로 포함되어 있다고 발표하였다[6], [7]. 일반적으로 행동인식을 위한 데이터 수집을 위해서 사용하는 샘플링 주파수는 센서에 따라 다른 값을 가지며, 가속도 센서의 경우에는 20~50Hz의 샘플링 주파수를 주로 사용하고 있다.

효율적인 행동인식을 수행하기 위해서는 센서로부터 수집된 데이터를 분석에 적합한 데이터로 변환하는 과정이 필요하다. 일반적으로 수집되는 데이터에서 불필요한 정보들을 제거하고 행동정보에 적합한 데이터를 추출하는 작업을 전처리 과정이라 한다. 전처리 과정에는 수집된 신호에서 불필요한 영역이나 신호에 포함된 노이즈를 제거하고 데이터의 정규화 또는 분석에 적합한 형태로 데이터 단위를 구분(Segmentation)하는 일련의 과정들을 포함한다.

2. 특징 추출 및 모델 학습

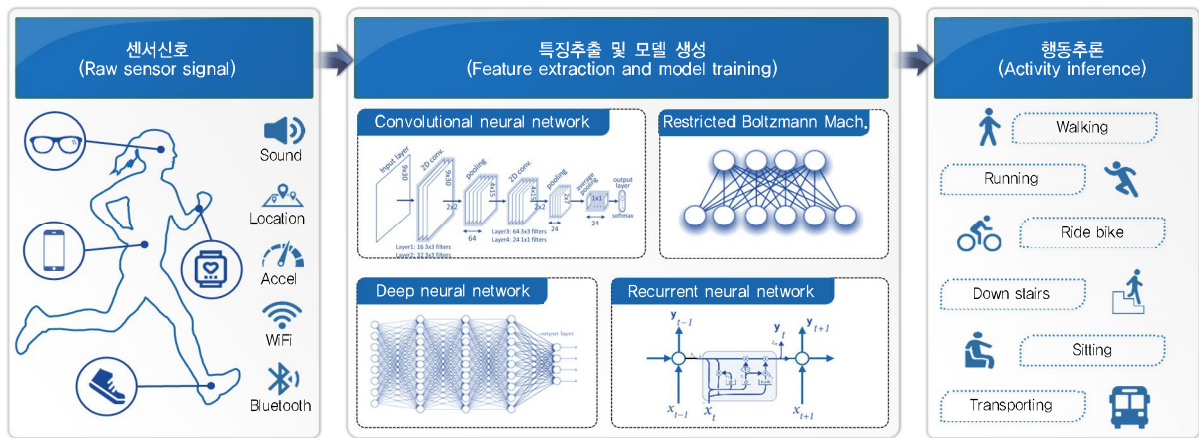
행동인식에서 가장 중요한 부분은 입력 신호에서 분석할 행동들에 적합한 특징(Feature) 정보들을 추출하는 것이다. 행동인식을 위해 일반적으로 사용되는 특징들은 크게 시간 영역(Time Domain)과 주파수 영역(Frequency Domain)의 특징으로 구분할 수 있으며[8], [9], 대표적으로 사용되는 특징들을 <표 1>에 정리하였

<표 1> 행동인식을 위한 일반적 특징[8], [9]

영역	특징	활용
시간	표준편차	일반적인 행동인식
	제곱평균	걸음 패턴
	평균값	앉은 자세와 서있는 자세
	신호크기영역	활동구간과 비활동 구간
	신호벡터영역	이동강도도(낙상 구분)
주파수	에너지	걷기, 달리기, 앉기, 서있기
	엔트로피	걷기, 달리기, 앉기, 서있기
	FFT peak	일반적인 행동인식

다. <표 1>과 같이 이들 개별 특징들은 각각 특정한 행동이나 상태에 대하여 다른 특징들과 구분되는 성능을 나타내므로 인식하고자 하는 행동의 특성에 적합하도록 다양한 특징들을 조합하여 사용한다.

일반적으로 동일하거나 유사한 행위의 반복적인 표현으로 정의되는 저수준 행동의 인식을 위한 학습에는 <표 1>의 특징들을 활용한 기계학습(Machine Learning) 방법을 사용한다. 기계학습이란 데이터를 컴퓨터에 입력하면 컴퓨터가 스스로 훈련하면서 패턴이나 특징을 찾아내서 비슷한 것들끼리 분류하고 판독하도록 학습하는 방법을 의미한다. 데이터를 어떻게 분류할 것인가에 대하여 이미 많은 기계학습 알고리즘들이 제안되고 활용되어왔다. 대표적으로 의사결정트리(Decision tree), 베이지안망, KNN(K-Nearest Neighbor), 서포트 벡터머신(SVM: Support Vector Machines), 인공신경망(ANN: Artificial Neural Network) 알고리즘을 들 수 있다. 각 알고리즘마다 장단점이 분명해 분석하려는 데이터의 속성이나 형태에 따라 서로 다른 알고리즘이 사용된다. 예를 들어, 의사결정트리는 다른 분류 알고리즘보다 계산 비용이 적고 이해하기 쉬운 모델을 생성할 수 있다는 장점이 있다. KNN은 입력되는 인스턴스와 모든 단일 학습 인스턴스를 비교하는 인스턴스 기반의 학습(IBM: Instance-Based Learning)을 사용하므로 계산상 많은 연산이 필요하지만 다른 알고리즘에 비하여 새로운 데이터에 유연하게 대응할 수 있는 장점이 있어서 많



(그림 2) 딥러닝을 활용한 행동인식 기술

이 사용되고 있다.

기존에 주로 사용된 이러한 기법들은 데이터로부터 유용한 시공간적 특징들을 추출할 때, 연구자가 직접 설계한 특징(Hand-Crafted Features)들을 이용하였다. 따라서 이러한 기법들은 추출할 특징들을 연구자가 직접 설계해야 하므로 상대적으로 많은 시간과 노력이 필요하고 특징들을 효율적으로 추출하기 위해서는 적용분야에 관련한 어느 정도의 전문 지식이 필요하다. 또한, 분석하려는 데이터의 종류와 성격에 따라 효율적인 특징들이 달라짐에도 불구하고, 미리 설계하여 고정된 특징들만 분석에 이용하기 때문에 새롭고 다양한 데이터에 대응하기 어렵고 일반화된 모델을 생성할 수 없다.

최근에는 이러한 단점들을 보완하기 위해 데이터로부터 직접 유용한 시공간적 특징들을 학습하는 딥러닝 방식이 주로 사용되고 있다. 딥러닝은 심층신경망(DNN: Deep Neural Networks)을 활용한 기계학습 방법이다. 기존의 행동인식을 위해 사용된 기계학습 알고리즘과 달리, 딥러닝 기술을 활용한 행동 인식은 모델 학습을 위한 특징들을 자동으로 추출하므로 보다 일반적인 모델 학습이 가능하다는 장점이 있다. (그림 2)에는 딥러닝을 이용한 행동인식의 간략한 프로세스를 나타내었다.

딥러닝 모델의 신경망은 여러 단계의 은닉 계층(Hidden Layer)을 포함하며, 각 은닉 계층들은 이전 계

층에서 가져온 출력의 지역화된 영역에서 컨볼루션 또는 다운 샘플링을 수행하며 네트워크를 구성한다. 이러한 구성적 특징을 가지는 딥러닝 네트워크는 입력 데이터의 로컬 영역 간 상관관계를 활용할 수 있기 때문에 주로 이미지 처리에 활용되어 왔다. 딥러닝을 사용한 행동인식에는 심층 신경회로망(DNN)이나 심층 합성곱 신경회로망(DCNN: Deep Convolutional Neural Networks)과 같은 모델이 주로 사용되고 있으며, 주로 입력되는 신호를 이미지 형태로 변환하여 사용한다. 센서 신호의 이미지 변환은 센서 신호의 입력 값을 이미지 형태로 직접 재구성하거나, 스펙트로그램(Spectrogram) 또는 신호의 세그먼트로부터 특징들을 추출한 후 이미지 형태로 변환하는 방식을 사용한다.

대표적인 예로써, [10]은 가속도 센서와 자이로 센서의 신호를 이용하여 2단계에 걸친 이미지 변환 방식을 제안하였다. 먼저 3축 자이로 센서의 신호와 3축 가속도 센서의 전체 가속신호와 선형 가속신호를 포함하는 9가지 신호를 서로 연관성을 가질 수 있도록 3가지 센서 신호를 x, y, z 축의 순서로 연결하여 2차원 신호 이미지로 생성하였다. 생성된 신호 이미지에 대하여 2차원 DFT(Discrete Fourier Transform) 크기를 계산하여 액티비티 이미지를 추출하고 DCNN 네트워크의 입력으로 사용하였다. 실험결과 DCNN 분류기는 기존의 SVM

분류기를 사용한 경우보다 행동인식에 향상된 성능을 보여주었다. 추가적으로 DCNN의 분류 결과인 확률 출력 값들의 차이가 명확하지 않은 경우 더 높은 성능을 위해 연구자가 직접 설계한 특징을 기반으로 하는 이진 SVM 분류기를 추가로 사용하여 성능을 향상하는 방안도 추가로 제시하였다. [11]에서는 3축 가속도 센서와 3축 자이로스코프에서 입력되는 데이터에 단시간 푸리에 변환(STFT: Short-Time Fourier Transform)을 적용하고 스펙트로그램으로 변환하여 이미지를 생성하는 방법을 제안하였다. 생성된 이미지는 CNN(Convolutional Neural Network)을 활용한 행동인식기의 입력으로 사용하였다.

스마트폰 기반의 행동인식 기술은 시간에 따른 연속적인 데이터를 처리를 필요로 한다. 사용자의 행동은 결국 시간에 따른 연속적인 데이터이므로 특정 시점의 데이터도 중요하지만, 시간에 따른 행동의 변화에 대한 분석이 중요하기 때문에 최근에는 순환 신경망(RNN: Recurrent Neural Network) 모델을 사용하는 방법도 제시되고 있다. 또한, 최근 들어 스마트폰에 내장된 여러 종류의 센서로부터 데이터와 더불어 다양한 환경 데이터를 활용하는 추세이므로 다중 데이터(Multi-modal)의 분석도 요구된다. 딥러닝은 이러한 여러 종류의 데이터들을 활용하여 현재 수행되는 사용자 행동에 대한 종합적인 판단을 내릴 수 있는 가장 적합한 방법으로 인식되고 있다.

3. 행동인식 관련 기술 이슈

최근 스마트폰 기반의 행동인식 기술은 다양한 연구를 통해 비약적인 발전을 하였으나, 여전히 해결해야 할 여러 가지 기술적 이슈들이 있다[12]. 먼저 모바일 플랫폼인 스마트폰이 기본적으로 갖는 문제로 온라인 배포 및 모바일 응용과 관련한 문제이다. 일부 온라인 학습을 이용한 딥러닝 기술이 제안되었으나 여전히 스마트폰에

적용하기에는 기술적인 한계가 있으며, 일반적으로 스마트폰 기반의 행동인식 모델은 대부분 원격 서버에서 오프라인 방식으로 학습되고 스마트폰에서는 학습된 모델만 사용하기 때문에 실시간으로 학습을 할 수 없다는 문제가 있다. 이를 해결하기 위해서는 모바일과 서버 간 통신 기술의 개선과 모바일 장치의 컴퓨팅 기능 향상이 필요하며, 스마트폰에서 실시간으로 학습할 수 있는 알고리즘적인 접근이 추가로 필요하다.

실시간 학습 문제와 더불어 추가로 스마트폰 기반의 행동인식 기술에서 고려해야 하는 문제는 알고리즘의 경량화 방안이다. 일반적으로 딥러닝 모델은 많은 컴퓨팅 리소스를 필요로하므로, 가용할 수 있는 컴퓨팅 파워나 메모리 리소스에 제약이 있는 스마트폰이나 웨어러블 장치에 사용하기에는 어느 정도 한계가 있다. 반면에 행동인식을 위해 얇은 구조를 가지는 신경망 네트워크 또는 기존의 기계학습 방법을 사용하였던 모델은 높은 성능을 언더라도 일반화된 모델을 생성하기 어렵다. 따라서 스마트폰에서의 효율적인 행동인식을 위해서는 보다 경량화된 딥러닝 모델의 개발이 필요하다.

마지막으로 알고리즘이나 모델의 정확도는 학습과정에 사용하는 데이터가 제공하는 행동의 범주와 데이터의 규모 및 데이터 수집에 참여한 사람들의 행동 특성에 영향을 받는다는 사실이다. 또한, 스마트폰 기반의 행동인식을 위하여 사용한 데이터가 실험실 기반의 제한된 환경에서 실제 외부 환경으로 옮겨지면 디바이스를 소지하는 위치나 방향에 따라 정확도가 떨어질 수 있다. 따라서 학습에 사용하는 데이터 셋의 효율적인 구축을 위해서는 이러한 상황을 반영하여야 한다.

지금까지 간략하게 살펴보았듯이 딥러닝은 행동인식에 관련하여 시공간 특징들을 자동으로 추출하고 심층적으로 구성함으로써 연구자의 지식을 기반으로 하는 기존 방식의 한계를 상당 부분 개선하였다. 그러나 스마트폰을 활용한 행동인식을 위해서는 여전히 플랫폼 환

경이나 알고리즘 측면에서 극복해야 할 문제가 많이 있음도 확인하였다. 특히 딥러닝 모델의 학습에는 사용하는 데이터의 종류와 데이터 수집에 참가하는 사람들의 특성 등에 많은 영향을 받으므로 데이터 셋의 연관성이 크다. 따라서 효율적인 학습을 위해 필수적인 데이터 셋의 구축 동향에 대하여 다음 장에서 보다 자세히 분석하였다.

III. 행동인식을 위한 데이터베이스 동향

행동인식을 위한 센서 기반의 데이터베이스는 2011년 이후 스마트폰의 센서 데이터만 사용한 10여 개의 데이터 셋을 포함해서 공개된 데이터 셋만 20여 개가 넘을 정도로 많이 구축되었다[12]. 특히, 최근에 공개되고 있는 일상생활의 행동 인식을 위한 데이터 셋은 대부분 스마트폰을 기반으로 하고 있으며, 기존의 저수준 행동뿐만 아니라 일상생활 중 낙상과 같은 돌발적인 상황의 데이터까지 포함하여 데이터베이스를 구축하는 사례가 점점 증가하고 있다.

행동인식을 위한 데이터베이스를 구축하는 방법은 크게 실험실 등 제한된 환경에서 수집한 데이터를 활용하여 데이터베이스를 구축하는 방법과 실제 생활에서 사용자의 자연스러운 행동 데이터를 수집하여 구축하는 방법으로 구분할 수 있다. 현재까지 구축된 대부분의 데이터 셋은 실험실 환경과 같이 제한된 환경에서 수집한 데이터를 사용하였으나 최근에 아마존과 UC Irvine(UCI)에서 실제 생활에서 사용자의 행동 데이터를 수집할 수 있는 새로운 방식을 제안하고 수집한 데이터 셋을 공개하였다[13], [14].

1. 제한된 환경의 데이터베이스

제한된 환경에서의 데이터 수집은 실험실 환경에서 연구자가 제공한 디바이스를 실험자 신체의 특정한 위치에 부착하여 데이터를 수집하는 것을 의미한다. 연구

자가 제공한 디바이스를 특정 위치에 부착한 실험자는 연구자가 사전에 정의한 스크립트에 따라서 특정 행동을 반복적으로 수행하게 되며, 행동과 행동 사이에는 정지 동작을 넣어서 동작을 구분하게 된다. 이와 같이 제한된 환경에서 행동 데이터를 수집하는 경우 수집할 수 있는 데이터의 종류와 양이 제한되며, 실험에 참가하는 사용자의 부자연스러운 행동 데이터를 수집할 수 있는 문제점이 있다. 따라서 제한된 환경에서 수집된 데이터를 기반으로 생성된 행동인식기를 실생활에 적용하는 경우 행동인식 성능이 저하되는 문제가 발생할 수 있다.

스마트폰의 센서 데이터를 사용하여 제한된 환경에서 행동인식 데이터를 수집한 대표적인 데이터베이스는 UCI 데이터 셋[15]이다. UCI 데이터 셋은 2012년 UCI에 의해서 공개되었으며, 30명의 지원자에게 대해서 걷기, 서기, 앉기, 눕기, 계단 오르기, 계단 내려가기 등의 6개 동작에 대한 데이터를 수집하였다. 데이터 수집을 위한 스마트폰으로 삼성 갤럭시 S2를 사용하였으며, 3축 가속도 신호 및 3축 각속도 신호를 50Hz로 샘플링하여 수집하였다. 수집된 개별 행동에 대한 데이터는 128개의 샘플 윈도우(약 2.56초) 단위로 구성하였으며, 윈도우를 사용한 각 데이터는 50%씩 겹치게 샘플링되었다.

전체 데이터의 샘플 수는 10,299개이며, 이 중 70%는 학습데이터로 사용되고 나머지 30%의 데이터들은 테스트 데이터로 사용된다. 데이터 셋을 공개한 논문에서는 각 윈도우로부터 타임 도메인과 주파수 도메인으로부터 추출한 17개의 특징(가속도 신호의 평균, 표준편차, 엔트로피 등)과 멀티클래스 SVM을 사용했을 때 89.3%의 정확도를 보여주었다. 데이터 셋의 공개 이후 UCI 데이터에 딥러닝을 적용한 다양한 방법들이 제안되었으며, CNN을 사용한 방법이 95%의 정확도로 가장 높은 성능을 나타내고 있다[16].

스마트폰 센서 기반의 다른 대표적인 행동 인식 데이터 셋은 WISDM(Wireless Sensor Data Mining) 데이터

셋[10]이다. WISDM 데이터 셋은 2010년 공개된 데이터 셋과 2012년 공개된 데이터 셋으로 구성되어 있다. WISDM 데이터 셋은 행동인식에 사용되는 여러 가지 신호 중 스마트폰의 3축 가속도 센서만 이용하여 20Hz로 샘플링하여 데이터를 수집하였으며, Nexus One, HTC의 Hero, Motorola의 Backflip 등 다양한 안드로이드 기반의 스마트폰을 사용하였다. 수집한 데이터는 일상생활의 6가지 저수준 행동(Walking, Jogging, Upstairs, Downstairs, Sitting, Standing)을 대상으로 하였으며, 일반적으로 사용하는 평균, 표준편차, Binned 분포 등을 포함하는 6개의 기본적인 특성을 활용하여 43개의 요약된 특징들을 생성하여 학습에 활용함과 동시에 개별 행동에 대하여 3축 가속도 센서의 원래 데이터(Raw data)를 함께 제공하고 있다. 수집된 데이터 셋의 검증에 위해 자바 기반의 오픈 소스 데이터 마이닝 프로그램인 WEKA(Waikato Environment for Knowledge Analysis)에서 제공하는 결정트리(J48), 로지스틱 회귀(Logistic Regression) 및 다층 신경망 3가지의 분류기법에 대하여 10-폴드 교차검증을 사용한 결과 J48은 85%, 로지스틱 회귀는 78%, 다층 신경망은 91.7%의 인식 성능을 가진다고 발표하였다. 이후 WISDM 데이터 셋은 CNN 및 RNN 등의 딥러닝을 적용한 다양한 행동 인식 방법들의 검증에 사용되고 있다.

최근 2017년에는 스마트폰 센서신호를 기반으로 하는 UniMiB SHAR 데이터 셋[17]이 공개되었다. UniMiB SHAR 데이터 셋은 삼성 갤럭시 넥서스 3를 데이터 수집을 위한 플랫폼으로 사용하였으며, 스마트폰을 실험자의 좌우 바지주머니에 한 번씩 넣고 3축 가속도 센서 데이터를 50Hz로 샘플링하여 수집하였다. 본 데이터 셋은 9개의 ADL(Activity of Daily Living)과 8종류의 낙상 행위 등 총 17개의 행동으로 구성되어 있으며, 30명의 실험자로부터 11,771개의 데이터를 수집하였다. 데이터를 공개한 논문에서는 KNN, SVM,

ANN, RF(Random Forest) 등의 방법을 기본 분류 성능으로 제시하고 있으며, 5-폴드 교차 검증을 통하여 성능을 실험한 결과 9개의 일상생활 행동을 구분하는 성능은 RF를 사용했을 때 약 88.41%의 성능을 보여주었다.

UniMiB SHAR 데이터 셋의 경우 일반화 성능을 평가하기 위하여 LOO(Leave-One-Out) 프로토콜을 사용한 성능을 추가로 제시하고 있다. LOO 프로토콜은 데이터 셋을 구성하는 데 참여한 N 명의 실험자 중 k 번째 실험자의 데이터를 제외한 나머지 모든 데이터를 학습 데이터로 사용하고, k 번째 실험자의 데이터를 테스트 데이터로 사용하여 성능을 평가한 후, 이 과정을 전체 참여자(N)에 대하여 반복하는 방법이다.

9개의 일상 생활행동을 구분함에 있어서 5-폴드 교차검증을 사용했을 때의 최고 성능은 88.41%였지만, LOO 프로토콜을 사용하여 실험하는 경우 최고 성능은 73.17%로 약 15% 이상의 성능저하가 발생하였다. 이러한 현상은 센서의 위치 및 행동을 제한하여 획득한 데이터로 학습한 분류 모델은 다양한 환경에서 다른 사람들의 행동 특성을 반영하지 못하여 인식 성능이 급격하게 떨어지는 것을 의미한다.

2. 실제 생활을 고려한 데이터베이스

앞에서 살펴본 제한된 데이터 환경에서 구축된 3가지 대표적인 데이터베이스의 경우 최신의 딥러닝 기술을 적용하면 10개 내외의 행동 범주를 90% 이상의 높은 성능으로 분류할 수 있다는 것을 확인하였다. 하지만 최근 UniMiB SHAR 데이터 셋의 실험 결과에서 알 수 있듯이 센서의 위치와 방향, 데이터를 수집하는 센서의 종류 등 데이터를 수집하는 환경과 실제 적용하는 환경이 달라지는 경우 급격한 성능 저하가 발생한다는 것을 확인하였다. 이러한 현상은 제한된 환경에서 구축된 데이터 셋을 이용하여 모델을 학습하는 방법이 가지는 한계이며, 이를 극복하고자 최근에 실생활에서 사용자의 행동

데이터를 자연스럽게 수집할 수 있는 방법에 대한 연구가 진행되고 있다.

최근 아마존과 UCI에서 일상생활에서 사용자의 행동 데이터를 자연스럽게 수집할 수 있는 방법을 제안하고, 제안한 방법으로 2015년부터 2016년간 수집된 데이터들을 이용하여 구축한 ExtraSensory 데이터 셋[18]을 공개하였다. ExtraSensory 데이터 셋은 사용자가 하는 다양한 일상생활에서 일어나는 행동에 대한 데이터를 자연스럽게 수집하기 위하여 센서 데이터를 수집하는 별도의 앱(ExtraSensory App)을 개발하였다. 데이터 수집을 위해 특정 종류의 스마트폰을 사용하였던 기존 방법과 달리, 실험자는 자신이 사용하는 스마트폰에 앱을 설치하고, 연구자가 제공하는 스마트 워치를 차고 일상생활을 하면서 자연스럽게 데이터를 수집하도록 설계되었다.

앱을 기반으로 데이터를 수집하면 기존의 연구자가 제공하던 기기를 사용하는 제한된 환경과 달리 사용자가 소유한 기기를 활용할 수 있으며, 이는 사용자들이 사용하는 기기 및 센서의 다양성을 반영할 수 있는 데이터 셋을 구축할 수 있는 장점이 있다. 또한, 연구자가 데이터 수집을 목적으로 스마트폰의 위치나 사용자가 별도로 수행해야 될 행동을 지정해주지 않기 때문에 실험자는 평소 일상생활을 수행하는 과정에서 자연스럽게 행동 데이터, 다양한 위치에서의 센서 데이터를 수집할 수 있게 된다.

따라서, 이러한 방식으로 구축하는 데이터 셋은 데이터를 수집하는 센서와 사용하는 기기의 다양성, 기기가 착용된 다양한 위치 및 많은 사람의 가진 고유한 행동 특성들이 모두 반영되어 있기 때문에, 이러한 형태의 데이터 셋에 기반하여 학습된 행동인식 방법들은 실제 생활에 적용하여도 급격한 성능 저하가 발생되지 않는다. ExtraSensory 데이터 셋의 또 다른 특징으로는 행동 데이터를 멀티 레이블로 태깅하였다는 점이다. 기존의 데이터 셋들의 경우, 샘플 하나로 한 가지 행동만 판단했

〈표 2〉 ExtraSensory 수집 데이터

센서	영역	데이터 수
Accelerometer	3-axis, 40Hz, 20s	308,306
Gyroscope	3-axis, 40Hz, 20s	291,883
Magnetometer	3-axis, 40Hz, 20s	282,527
Watch 가속도	3-axis, 25Hz, 20s	210,716
Watch compass	Heading angle	126,781
Location (오프라인 계산)	Long-lat-alt	273,737
Location (스마트폰 계산)	위치 정보(1분에 한번)	263,899
Audio	프레임 당 13 MFCC* (96ms 프레임, 50% 겹침)	302,177
Audio magnitude (정규화된 최대값)	1분에 한번	303,877
Phone state (앱, 배터리, WiFi등)	1분에 한번	308,320
기타센서	1분에 한번	308,312
Core	-	176,941

* MFCC: Mel-Frequency Cepstral Coefficients

기 때문에, 일상생활에서 나타나는 복합적이고 다양한 행동의 특성을 반영하지 못하는 단점이 있었다. 예를 들어, ‘걷기’의 경우 기존 데이터 셋의 경우 단순하게 걷는다는 행위 자체에 대한 저수준의 행동만을 나타내지만, ExtraSensory에서는 ‘커피를 들고 걷다’, ‘짐을 들고 걷다’, ‘친구와 걷다’ 등의 부가적인 정보들이 같이 태깅할 수 있게 함으로써 보다 상위수준의 행동인식에 접근할 수 있게 하였다. 이러한 상위 수준의 행동인식을 위해서 ExtraSensory 앱에서는 스마트폰의 가속도 신호, 자이로스코프, 스마트워치의 가속도 신호, 위치, 주변 소리, 핸드폰 상태 등 6개의 필수적인 센서 정보와 다른 부가적인 센서 정보를 수집하였으며, 상세한 내용은 〈표 2〉와 같다.

ExtraSensory는 1분마다 20초 동안의 윈도우를 사용하여 표에서 정리된 센서 데이터들을 수집하였으며, 60명으로부터 총 308,320개의 분 단위로 수집된 데이터로 구성된다. 실생활에서 데이터를 수집함에 있어서 가장 큰 문제는 사용자의 행동에 대한 정확한 Ground Truth를 얻는 것으로, ExtraSensory에서는 다양한 메커니즘

을 사용하여 수집된 데이터의 라벨링 과정에 사용자의 적극적인 참여를 유도하고, 사용자가 피드백을 통해 태깅한 데이터만을 Ground Truth 데이터로 사용하였다.

ExtraSensory 연구에서는 수집하는 센서 데이터에 각각 개별적인 인식기를 사용한 경우와 다양한 알고리즘을 복합적으로 사용하여 센서 정보를 융합한 경우를 5-폴드 교차 검증을 통하여 성능을 측정했다. 그 결과 25개 콘텍스트 범주에 대해, 센서 정보들을 융합했을 때 단일 센서 정보를 사용하는 경우보다 10~15% 이상 향상된 약 87%의 정확도를 보였다. 또한, 실제 생활 환경에서 발생할 수 있는 여러 가지 다양성을 갖는 데이터 셋이므로 LOO 프로토콜을 사용하여 성능을 측정했을 때에도 5-폴드 교차 검증과 거의 유사한 86%의 정확도를 나타내었다.

본 절에서 살펴보았듯이 기존의 실험실 환경에서 수집된 제한적인 데이터 셋은 일부 저수준의 행동인식에는 효율적인 인식 성능을 도출할 수 있었다. 그러나 행동인식 기술이 실제로 적용되는 일상생활에서는 인식 성능이 현저하게 감소함을 알 수 있었다. 이러한 현상은 결국 데이터 수집 환경과 방법에서 실제 생활 환경을 고려하지 않아서 발생하는 문제로서, 최근 이러한 문제를 극복하기 위한 방법으로 실제 생활에서 데이터를 수집하는 기술적 접근이 진행되고 있음을 살펴보았다. 행동인식 기술의 궁극적인 목표는 다양한 환경에서 사용자들의 행동을 인식하는 것임을 상기해보면, 최근의 실제 생활 환경을 고려한 데이터 수집과 이러한 데이터를 이용한 데이터베이스 구축 방향은 앞으로도 지속될 것이라 판단된다.

IV. 결론

사용자 행동인식(HAR: Human Activity Recognition) 기술은 다양한 센서를 활용하여 사람의 모션이나 제스처와 관련된 정보를 수집하고 해석하여 행동을 인

식하는 기술로써, 다른 사람과의 효과적인 소통을 포함한 다양한 서비스 제공을 위해 필요한 기술이다. 최근 스마트폰 기술의 비약적인 발전과 함께 일상생활에서 스마트폰의 활용도가 높아지면서 스마트폰을 활용한 다양한 행동인식 기술이 개발되고 있다.

본고에서는 스마트폰 기반 사용자 행동인식 기술의 전반적인 동향과 향후 필요한 기술적 이슈들을 살펴보았다. 스마트폰에서의 효율적인 행동인식을 위해서는 보다 경량화된 모델의 개발과 함께 학습된 모델의 온라인 배포 및 다양한 응용과 관련하여 해결하여야 할 문제들이 산재함을 알 수 있었다. 또한, 효율적인 행동인식을 위해 모델 학습에 사용되는 양질의 데이터 확보가 필수적이므로 행동 인식 학습에 활용되는 다양한 데이터 셋 구축을 위한 최근 개발 동향을 함께 분석하였다.

행동 인식 데이터 셋은 제한된 환경에서 수집되는 저수준의 기본 데이터 셋 구축에서 사용자의 일상생활에서 자연스럽게 데이터를 수집하고 보다 복합적인 데이터들을 수집하는 방향으로 전환되고 있음을 확인하였다. 이러한 추세는 다양한 데이터들이 혼재되어 사용되는 빅데이터 환경에서 대규모의 데이터 분석을 위한 자연스러운 흐름으로 분석된다. 아울러 아직까지 사용자 행동인식에 필요한 공개 데이터 셋이 거의 전무한 국내에서도 연구 개발에 필요한 데이터들을 확보하기 위해 전략적인 노력이 시급한 시점이라고 판단된다.

약어 정리

ADL	Activity of Daily Living
ANN	Artificial Neural Network
CFS	Correlation Feature Selection
CNN	Convolutional Neural Network
DCNN	Deep Convolutional Neural Networks
DFT	Discrete Fourier Transform
DNN	Deep Neural Networks
FFT	Fast Fourier Transform
HAR	Human Activity Recognition

HCI	Human-Computer Interaction
HRI	Human-Robot Interaction
IBL	Instance-Based Learning
KNN	K-Nearest Neighbor
LOO	Leave-One-Out
MFCC	Mel-Frequency Cepstral Coefficients
RF	Random Forest
RFID	Radio Frequency Identification
RNN	Recurrent Neural Network
STFT	Short-Time Fourier Transform
SVM	Support Vector Machines
WEKA	Waikato Environment for Knowledge Analysis
WISDM	Wireless Sensor Data Mining

참고문헌

- [1] H. Osman, “차세대 스마트폰 시장 동력 기술은 5G·AI·행동 인식” 가트너,” CIO, Accessed Mar. 29, 2018. <http://www.ciokorea.com/news/34639>
- [2] D. Cook, K.D. Feuz, and N.C. Krishnan, “Transfer Learning for Activity Recognition: A Survey,” *Knowl. Inform. Syst.*, vol. 36, no. 3, Sept. 2013, pp. 537–556.
- [3] H. Martín, A.M. Bernardos, J. Iglesias, and J.R. Casar, “Activity Logging Using Lightweight Classification Techniques in Mobile Devices,” *Personal Ubiquitous Comput.*, vol. 17, no. 4, Apr. 2013, pp. 675–695.
- [4] L. Bao and S. Intille, “Activity Recognition from User-Annotated Acceleration Data,” in *Pervasive Computing. Pervasive 2004. Lecture Notes in Computer Science, vol 3001*, Berlin, Heidelberg: Springer, 2014.
- [5] O.C. Ann and L.B. Theng, “Human Activity Recognition: a Review,” *IEEE Int. Conf. Contr. Syst. Comput. Eng.*, Batu Ferringhi, Malaysia, Nov. 28–30, 2014, pp. 389–393.
- [6] M. Shoaib, S. Bosch, O.D. Incel, H. Scholten, and P.J.M. Havinga, “A Survey of Online Activity Recognition Using Mobile Phones,” *Sensors*, vol. 15, no. 1, 2015, pp. 2059–2085.
- [7] M.A.A. de la Concepción, L.M. Soria Morillo, L. Gonzalez-Abril, and J.A. Ortega Ramirez, “Discrete Techniques Applied to Low-Energy Mobile Human Activity Recognition. A New Approach,” *Expert Syst. Applicat.*, vol. 41, no. 14, Oct. 2014, pp. 6138–6146.
- [8] S.D. Bersch, D. Azzi, R. Khusainov, I.E. Achumba, and J. Ries, “Sensor Data Acquisition and Processing Parameters for Human Activity Classification,” *Sensors*, vol. 14, no. 3, 2014, pp. 4239–4270.
- [9] S. González, J. Sedano, J.R. Villar, E. Corchado, Á. Herrero, and B. Baruaque, “Features and Models for Human Activity Recognition,” *Neurocomput.*, vol. 167, Nov. 2015, pp. 52–60.
- [10] J.R. Kwapisz, G.M. Weiss, and S.A. Moore, “Activity Recognition using Cell Phone Accelerometers,” *Sensor KDD*, Washington, DC, USA, July 25, 2010, pp. 1–9.
- [11] D. Ravi, C. Wong, B. Lo, and G.Z. Yang, “Deep Learning for Human Activity Recognition: A Resource Efficient Implementation on Low-Power Devices,” *IEEE Int. Conf. Wearable Implantable BSN*, San Francisco, CA, USA, June 14–17, 2016, pp.71–76.
- [12] J. Wang, Y. Chen, S. Hao, X. Peng, and L. Hu, “Deep Learning for Sensor-Based Activity Recognition: A Survey,” *Pattern, Recogn. Lett.*, Feb. 21, 2018.
- [13] Y. Vaizman, N. Weibel, and G. Lanckriet, “Context Recognition In-the-Wild: Unified Model for Multi-modal Sensors and Multi-label Classification,” *Proc. Interactive Mobile Wearable Ubiquitous, Technol.*, vol. 1, no. 4, Dec. 2017, pp. 168:1–168:22.
- [14] Y. Vaizman, K. Ellis, G. Lanckriet, and N. Weibel, “ExtraSensory App: Data Collection In-the-Wild with Rich User Interface to Self-Report Behavior,” *Proc. CHI Conf. Human Factors Comput. Syst.*, Montreal, Canada, Apr. 21–26, 2018, pp. 1–12.
- [15] D. Anguita, A. Ghio, L. Oneto, X. Parra, and J.L. Reyes-Ortiz, “Human Activity Recognition on Smartphones Using a Multiclass Hardware-Friendly Support Vector Machine,” in *Proc. of IWAAL, Lecture Notes in Computer Science, vol 7657*, Berlin, Heidelberg: Springer, 2012.
- [16] W. Jiang and Z. Yin, “Human Activity Recognition Using Wearable Sensors by Deep Convolutional Neural Networks,” *Proc. ACM Int. Conf. Multimedia*, Brisbane, Australia, Oct. 26–30, 2015, pp. 1307–1310.
- [17] D. Micucci, M. Mobilio, and P. Napolitano, “UniMiB SHAR: A Dataset for Human Activity Recognition Using Acceleration Data from Smartphones,” *Appl. Sci.*, vol. 7, no. 10, 2017, pp. 1101:1–1101:19.
- [18] Y. Vaizman, K. Ellis, and G. Lanckriet, “Recognizing Detailed Human Context In-the-Wild from Smartphones and Smartwatches,” *IEEE Pervasive Comput.*, vol. 16, no. 4, 2017, pp. 62–74.