

## 웨어러블 센서를 이용한 사건인지 기반 일상 활동 예측\*

이충연<sup>1</sup>○, 곽동현<sup>2</sup>, 이범진<sup>1</sup>, 장병탁<sup>1</sup><sup>1</sup>서울대학교 컴퓨터공학부, <sup>2</sup>서울대학교 뇌과학협동과정

{cylee, dhkwak, bjlee, btzhang}@bi.snu.ac.kr

## Event Cognition-based Daily Activity Prediction From Wearable Sensors

Chung-Yeon Lee<sup>1</sup>○, Dong-Hyun Kwak<sup>2</sup>, Beom-Jin Lee<sup>1</sup>, Byoung-Tak Zhang<sup>1</sup><sup>1</sup>Department of Computer Science and Engineering, Seoul National University,<sup>2</sup>Interdisciplinary Programs in Neuroscience, Seoul National University

## 요약

웨어러블 장치와 스마트폰을 이용해 사용자의 일상 생활에서 일인칭 시점의 멀티 센서 데이터를 수집하고, 각각의 데이터를 통해 장소, 장면, 행동의 사건 요소를 인지하여, 최종적으로 사용자의 일상 활동을 예측하였다. 기존 연구들이 센서로부터 획득한 신호 데이터의 패턴과 일상 활동 사이의 직접적인 상관관계를 탐색한 반면, 제안하는 방법은 각각의 사건을 구성하는 요소를 먼저 파악하기 위해 GPS 데이터로부터 이벤트가 발생하는 장소를, 이미지 데이터로부터 활동 배경이 되는 장면을, 그리고 가속도와 자이로 센서 데이터로부터 사람의 신체 움직임을 각각 인식하고, 이 정보들이 조합되는 규칙을 학습하여 일상 활동을 예측한다. 스마트 안경을 착용하고 생활하면서 수집한 멀티센서 데이터로부터 19개의 장소, 32개의 장면, 7개의 움직임을 인식하고, 최종적으로 35개의 일상 활동을 84.91%의 정확도로 예측한 결과를 보임으로써 센서 데이터를 직접 사용하여 분류한 결과보다 심볼릭한 사건 요소 정보의 조합 규칙을 학습한 것보다 높은 정확도를 가질 수 있다는 사실을 확인할 수 있었다.

## 1. 서론

실제 환경에서 사람의 일상적인 활동을 자동으로 인지하고 예측하는 기술은 스마트 비서(assistant), 홈 로봇(home robot), 스마트 헬스케어(healthcare) 등 최근 각광받고 있는 스마트 머신 응용 분야에서 개인 서비스를 하기 위해 필요한 핵심 기술이다. 최근에는 GoPro나 Narrative Clip과 같은 웨어러블 카메라나 MS Band, Fitbit과 같은 스마트워치(Smart watch) 등이 출시되어 새로운 형태의 인간 활동 데이터 습득이 가능해지고 있다. 이러한 웨어러블 장치를 통해 획득되는 일인칭 시점의 센서 데이터는 사용자의 일상 활동을 객관적으로 반영하는 명시적인 정보를 제공할 수 있다는 장점이 있다[1]. 실제로 최근 헬스케어나 피트니스(fitness) 분야에서 사용자 데이터 수집이 활발하게 이루어지고 있으며, 기억 증강, 식이조절 평가, 자폐증 환자 보조, 여행 도우미, 신체 활동 인지 등에서 사용자 데이터 확보를 위해, 그리고 최근에는 사람의 행동을 스스로 관찰하여 인지 모델의 개념 메모리를 정의하도록 하는 인지 로봇 연구에서도 웨어러블 장치가 핵심적으로 사용되고 있다[2].

센서 데이터를 이용하여 사람의 일상 활동 패턴을 분석하고 인지하는 연구 또한 활발히 연구되는 분야이다. Biagioni와 Krumm [3]은 30명의 피험자가 46일간 수집한 GPS 센서 데이터로부터 개인의 특징적인 이동 패턴을 추출하는 알고리즘을 제시하였고, Sun 등 [4]은 Latent

Dirichlet Allocation 기법을 이용하여 다중 위치센서와 음향 데이터로부터 일상 활동을 예측할 수 있음을 보였다. Eagle과 Pentland [5]는 개인의 반복적인 일과를 구성하는 패턴을 추출하여 Eigenbehaviors라고 명시하였고, 이들의 가중치합을 이용하여 사람의 일상을 79%의 확률로 예측한 결과를 제시하였다. Clarkson [6]은 자신의 실제 일상에서 웨어러블 장치를 이용하여 100일간 수집한 센서 데이터들의 패턴들을 탐색하고 일과를 예측하였다.

이처럼 웨어러블 센서를 이용하여 일상을 예측하는 대다수의 연구들은 센서로부터 획득한 신호 데이터의 패턴과 일상을 구성하는 각각의 일과 사이의 상관관계를 찾는 것에 집중하였다. 하지만 일상에서의 인간 활동은 하나의 제목으로 표현하기 어려운 다수의 이벤트 집합이고 서술 가능한(narrative) 일종의 이야기(story)이다. 따라서 사람의 일상 활동을 학습하고 예측하기 위해서는 각각의 사건을 구성하는 요소를 먼저 파악하여야 한다.

본고에서는 다양한 기계학습 기법들을 이용하여 웨어러블 센서 데이터로부터 사건 요소들을 인지하고, 일상을 예측하는 방법에 대해 기술한다. 제안하는 방법은 GPS 데이터로부터 이벤트가 발생하는 장소를, 이미지 데이터로부터 활동 배경이 되는 장면을, 그리고 가속도와 자이로 센서 데이터로부터 사람의 신체 움직임을 각각 인식하고, 이 정보들이 조합되는 규칙을 학습하여 일상 활동을 예측한다. 제안하는 방법을 검증하기 위해 실험 참가자가 14일 동안 스마트 안경을 착용하고 생활하면서 수집한 멀티센서 데이터로부터 19개의 장소, 32개의 장면, 7개의 움직임을 인식하고, 최종적으로 35개의 일상 활동을 84.91%의 정확도로 예측한 결과를 제시한다.

\* 이 논문은 2015년도 정부(미래창조과학부)의 재원으로 정보통신기술진흥센터의 지원(R0126-15-1072-SW스타랩, 10035348-mLife, 10044009-HRI.MESSI)을 받아 수행된 연구임.

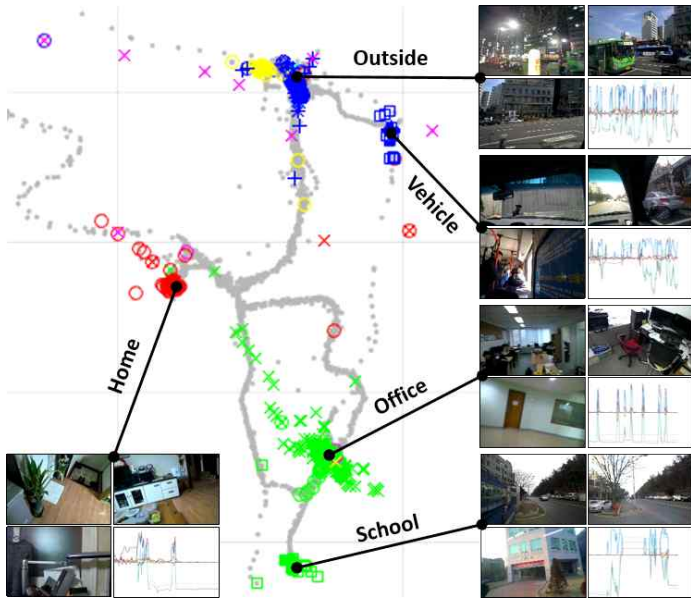


그림 1. 수집된 웨어러블 센서 데이터 예시

2. 연구 내용

2.1 웨어러블 센서 데이터 수집

실험 참가자가 웨어러블 장치를 착용하고 14일 동안 일상에서의 센서 데이터를 수집한다[7]. 수집된 데이터는 약 32만장의 이미지, 약 12만개의 GPS 좌표, 그리고 약 93만개의 3축 가속도/자이로/지자기 센서 데이터로 구성된다. 그림 1은 수집된 웨어러블 센서 데이터의 예이다. 이렇게 수집된 센서 데이터는 각각 19개 클래스의 장소(빌딩 단위의 거시적 장소), 32개의 장면(시각적 정보들로 표현되는 방 단위의 장소), 그리고 걷기, 앉기 등 7개의 신체 움직임을 반영한다. 데이터 수집시 실험 참가자의 활동 내역은 자체적으로 개발한 스마트폰 로깅 프로그램과 위치기반서비스(Location-based service, LBS) API를 연동하여 사용하였다. 또한 실험 참가자가 일차적으로 데이터를 스크리닝하여 사생활 침해의 우려가 있는 부분은 자율적으로 제거할 수 있도록 하였다.

2.2 사건 요소 인지 방법

수집된 센서 데이터로부터 그림 3과 같이 일상 활동에서 발생하는 이벤트의 구성 요소를 인지하여 추출한다.

**장소 인지:** 장소 인지를 위한 좌표 데이터는 인식 가능 범위에 따라 GPS, 적외선, 초음파 센서 등을 통해 구할 수 있으며, 로봇의 경우 SLAM 기법을 통해 구축된 지도 정보와 주행거리계(Odometer)를 이용하여 구할 수 있다. 본고에서는 기본적으로 GPS 데이터를 사용하고, 음영지역에서 누락된 좌표는 통신사의 기지국 정보를 통해 보정한다. 보정된 좌표 데이터는 그림 1과 같이 지도상에 나타낼 수 있다. 측정된 좌표로부터 장소를 인지하기 위해 해당 위치에서 사용자가 로깅한 LBS 정보를 이용하여 논리적 위치 레이블을 맵핑한다. 최종적으로 위도와 경도, 그리고 19개 클래스의 위치 레이블로 구성된 장소 데이터를 결정트리(Decision tree)로 학습하여 분류한다.

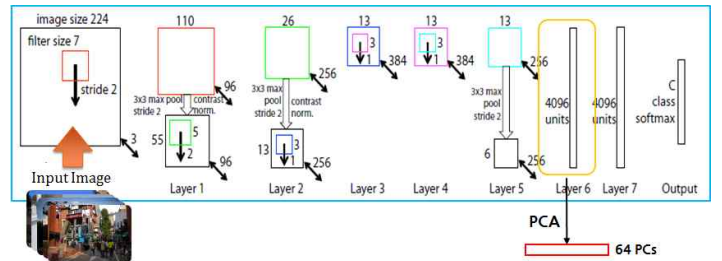


그림 2. 장면 특징 벡터 추출을 위한 CNN 구조

**장면 인지:** 웨어러블 카메라를 통해 획득한 이미지는 사람이 일상에서 활동하는 시각적 배경 정보를 가진다. 이 정보는 그 사람이 다루는 물체, 만나는 사람, 먹는 음식 등 다양한 맥락을 유추 가능하지만, 본고에서는 우선 사용자의 활동 배경이 되는 방 단위의 장면(scene)을 인지하여 활용한다. 각 이미지를 표현하는 확률적 특징 벡터를 추출하기 위해 ImageNet DB (ILSVRC)를 학습시킨 딥 컨볼루션 신경망(Deep convolutional neural networks, CNN)을 이용한다. 이때 계산에 소요되는 시간을 단축하기 위해 CUDA GPGPU 병렬 처리 컴퓨팅이 가능한 MatConvNet을 사용한다[8]. 그림 2는 CNN의 구조를 나타내며, 본고에서는 최종적으로 CNN의 6번째 레이어의 4096차원 값을 주성분분석(PCA)을 이용하여 64차원으로 축소하고, 이를 장면 인지를 위한 특징 벡터로 사용한다. 최종적으로 추출된 특징 벡터와 32개의 장면 레이블로 구성된 데이터를 SVM으로 학습하여 분류한다.

**행동인지:** 보행, 정지, 착석 등 사람의 움직임을 인지하기 위해 먼저 가속도/자이로/지자기 센서로부터 획득한 데이터로부터 Delta coefficient (DC), Shifted DC, 그리고 Signal magnitude area (SMA)를 계산하고, 이를 Random Forest로 학습하여 총 7개의 신체 행동을 분류한다.

DC는 주로 음성신호처리에서 시각적 정보를 표현하기 위해 쓰이던 방법으로 특정 구간에서의 신호를 시간에 대하여 미분한다. 이때 계산의 효율성을 위해 식 1처럼 근방의 몇 개의 점만 가지고 근사적으로 계산한다.

$$\Delta g(t) = \frac{\sum_{d=-D}^D dg(t+d)}{\sum_{d=-D}^D d^2} \tag{1}$$

SDC는 DC를 한 방향으로 전이시킨 값이다. 즉, 현재 프레임을 기준으로 앞, 뒤(과거와 미래) 프레임을 샘플링한 뒤 DC를 구해 현재 프레임의 특징벡터에 붙인다. SDC는 현재 특징 벡터 근처의 기울기가 시간에 따라 어떻게 변화하는지 수치적으로 나타내며 식 2와 같이 계산한다.

$$\Delta b(t+iP) = \frac{\sum_{d=-D}^D db(t+iP+d)}{\sum_{d=-D}^D d^2} \tag{2}$$

SMA는 식 3과 같이 3축 가속도 센서로부터 추출한 세 종류의 가속도 값을 가지고 한 윈도우 내에서 절대값의 합의 평균을 나타낸다.  $x[n]$ 과  $y[n]$ ,  $z[n]$ 은 각 축의  $n$ 번째 값을 나타내고  $N$ 은 SMA가 계산되는 윈도우 구간이다.

$$SMA = \frac{1}{N} \times \left( \sum_{n=1}^N (|x[n]| + |y[n]| + |z[n]|) \right) \tag{3}$$

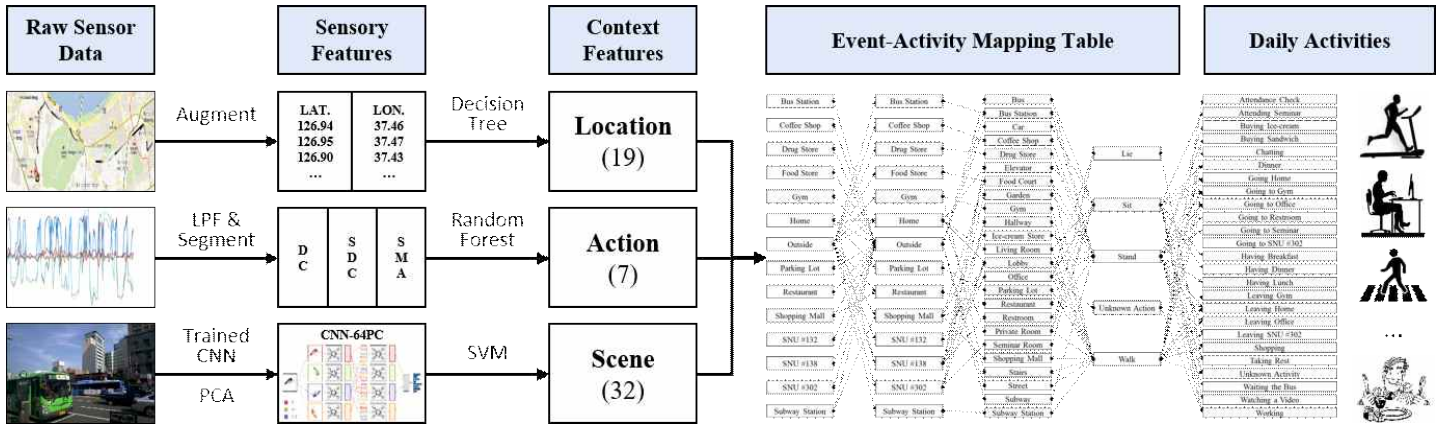


그림 3. 사건 인지 및 일상 활동 예측 프로세스 흐름도

### 2.3 일상 활동 예측 방법

위에서 구한 사건 요소들은 그림 3에서 볼 수 있듯이 19개의 장소(Location), 32개의 장면(Scene), 7개의 행동(Action)을 나타내는 심볼릭 정보들로 구성된다. 일상 활동에 대한 사건 요소들의 조합 규칙을 학습한 의사결정 테이블(Decision table)을 이용하여 최종적으로 사용자의 활동을 예측한다. 이때 현재 정보와 이전 상태에서의 정보를 결합하여 사용한다. 즉, 하나의 규칙을 구성하는 사건 요소들은 이전 장소, 현재 장소, 이전 장면, 현재 장면, 이전 행동, 현재 행동의 결합으로 표현된다.

### 3. 실험 결과

실험을 위해 수집된 전체 데이터 중 휴일을 제외한 나머지 10일 동안의 데이터를 8일의 학습 데이터와 2일의 검정 데이터로 나누었다. 이때 각 날짜에 포함된 사건 요소들의 조합이 편향되지 않도록 유사한 분포를 가지는 날짜의 데이터를 찾아 검정 데이터로 사용하였다.

먼저 각 사건 요소를 분류한 결과, 그림 4의 (a)와 같이 장소 인식은 90.75%, 장면 인식은 73.89%, 그리고 행동 인식은 92.25%의 정확도를 보였다. 그리고 이 결과를 이용해 일상 활동을 분류한 결과는 그림 4의 (b)와 같다. 세 가지 사건 요소들을 이용하여 분류한 결과는 84.91%이며, 비교를 위해 GPS, 이미지, 가속도 센서 데이터를 직접 이용하여 분류한 결과 각각 이보다 낮은 81.9%, 81.21%, 59.05%의 정확도를 보였다.

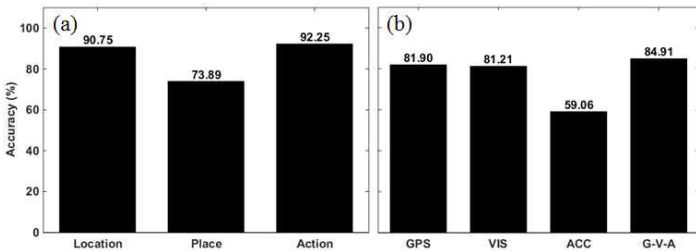


그림 4. 사건 인지 및 일상 활동 예측 결과

### 4. 논의 및 결론

본고에서는 웨어러블 장치와 스마트폰을 통해 사용자의 일상 생활에서 일인칭 시점의 멀티 센서 데이터를 수집하고, 각각의 데이터를 통해 장소, 장면, 행동의 사건

요소를 인지하여, 최종적으로 사용자의 일상 활동을 예측하는 프레임워크를 제시하였다. 실험 결과를 통해 센서 데이터를 직접 사용하여 분류한 결과보다 심볼릭(symbolic)한 사건 요소 정보의 조합 규칙을 학습한 것이 보다 높은 정확도를 가질 수 있다는 사실을 확인할 수 있었다. 단, 본 실험에서는 한 사람의 일상 활동만을 예측하였기 때문에 향후 다른 사람의 데이터를 이용한 추가 검증이 필요하며, 또한 서로 다른 사람의 유사한 일상 활동 학습 패턴이 전이되어 예측 가능한지에 대한 여부도 주요한 실험 이슈가 될 것으로 판단된다. 한편 최종 활동 예측을 위해 의사결정 테이블이 아닌 인공지능망 기법을 사용해볼 수 있을 것으로 기대된다.

### 참고 문헌

- [1] D. Castro, S. Hickson, V. Bettadapura, E. Thomaz, G. Abowd, H. Chirstensen, and I. Essa, "Predicting daily activities from egocentric images using deep learning", *Proc. International Symposium on Wearable Computers (ISWC-2015)*, pp. 75-82, 2015.
- [2] K. Ramirez-Amaro, M. Beetz, and G. Cheng, "Transferring skills to humanoid robots by extracting semantic representations from observations of human activities", *Artificial Intelligence*, 2015.
- [3] J. Biagioni and J. Krumm, "Days of our lives: Assessing day similarity from location traces", In *User Modeling, Adaptation, and Personalization*, pp. 89-101, Springer, 2013.
- [4] F.-T. Sun, Y.-T. Yeh, H.-T. Cheng, C. Kuo, and M.L. Griss, "Nonparametric discovery of human routines from sensor data", *Proc. Pervasive Computing and Communications (PerCom-2014)*, pp. 11-19, 2014.
- [5] N. Eagle, and A.S. Pentland, "Eigenbehaviors: Identifying structure in routine", *Behavioral Ecology and Sociobiology*, Vol. 63, No. 7, pp. 1057-1066, 2009.
- [6] B.P. Clarkson, "Life patterns: structure from wearable sensors", Ph.D. Thesis, MIT, Cambridge, MA, 2005.
- [7] 이충연, 곽동현, 곽하늬, 장병탁, "웨어러블 센싱 음향-영상 라이프로그 학습 기반 사용자 활동 인식", *향국컴퓨터종합학술대회(KCC)*, pp. 921-923, 2015.
- [8] A. Vedaldi, K. Lenc, "MatConvNet—convolutional neural networks for MATLAB," arXiv:1412.4564, 2014.