ISSN 2288-4866 (Print)
ISSN 2288-4882 (Online)
http://www.jiisonline.org
J Intell Inform Syst 2019 March: 25(1): 163~177 http://dx.doi.org/10.13088/jiis.2019.25.1.163

# 스마트폰 다종 데이터를 활용한 딥러닝 기반의 사용자 동행 상태 인식*

**김길호**
(주)수아랩
(*kim.kilho@sualab.com*)

**최상우**
(주)쿠팡 추천팀
(*sangwoochoi@coupang.com*)

**채문정**
서울대학교 산업공학과·
서울대학교 산업시스템혁신연구소
(*jjamjung@snu.ac.kr*)

**박희웅**
서울대학교 산업공학과·
서울대학교 산업시스템혁신연구소
(*hee188@snu.ac.kr*)

**이재홍**
카카오모빌리티 데이터랩
(*jaehong.lee@snu.ac.kr*)

**박종헌**
서울대학교 산업공학과·
서울대학교 산업시스템혁신연구소
(*jonghun@snu.ac.kr*)

스마트폰이 널리 보급되고 현대인들의 생활 속에 깊이 자리 잡으면서, 스마트폰에서 수집된 다종 데이터를 바탕으로 사용자 개인의 행동을 인식하고자 하는 연구가 활발히 진행되고 있다. 그러나 타인과의 상호작용 행동 인식에 대한 연구는 아직까지 상대적으로 미진하였다. 기존 상호작용 행동 인식 연구에서는 오디오, 블루투스, 와이파이 등의 데이터를 사용하였으나, 이들은 사용자 사생활 침해 가능성이 높으며 단시간 내에 충분한 양의 데이터를 수집하기 어렵다는 한계가 있다. 반면 가속도, 자기장, 자이로스코프 등의 물리 센서의 경우 사생활 침해 가능성이 낮으며 단시간 내에 충분한 양의 데이터를 수집할 수 있다. 본 연구에서는 이러한 점에 주목하여, 스마트폰 상의 다종 물리 센서 데이터만을 활용, 딥러닝 모델에 기반을 둔 사용자의 동행 상태 인식 방법론을 제안한다. 사용자의 동행 여부 및 대화 여부를 분류하는 동행 상태 분류 모델은 컨볼루션 신경망과 장단기 기억 순환 신경망이 혼합된 구조를 지닌다. 먼저 스마트폰의 다종 물리 센서에서 수집한 데이터에 존재하는 타임 스태프의 차이를 상쇄하고, 정규화를 수행하여 시간에 따른 시퀀스 데이터 형태로 변환함으로써 동행 상태 분류 모델의 입력 데이터를 생성한다. 이는 컨볼루션 신경망에 입력되며, 데이터의 시간적 국부 의존성이 반영된 요인 지도를 출력한다. 장단기 기억 순환 신경망은 요인 지도를 입력받아 시간에 따른 순차적 연관 관계를 학습하며, 동행 상태 분류를 위한 요인을 추출하고 소프트맥스 분류기에서 이에 기반한 최종적인 분류를 수행한다. 자체 제작한 스마트폰 애플리케이션을 배포하여 실험 데이터를 수집하였으며, 이를 활용하여 제안한 방법론을 평가하였다. 최적의 파라미터를 설정하여 동행 상태 분류 모델을 학습하고 평가한 결과, 동행 여부와 대화 여부를 각각 98.74%, 98.83%의 높은 정확도로 분류하였다.

**주제어** : 사용자 행동 인식, 그룹 상호작용, 스마트폰 물리 센서, 컨볼루션 신경망, 장단기 기억 순환 신경망

논문접수일 : 2018년 6월 25일 논문수정일 : 2019년 2월 12일 게재확정일 : 2019년 3월 11일
원고유형 : 일반논문 교신저자 : 최상우

* 본 연구는 삼성전자의 지원을 받아 수행한 결과임

## 1. 서론

스마트폰이 현대인들의 생활 속에 깊이 자리 잡으면서, 스마트폰에서 수집된 데이터를 바탕으로 스마트폰 사용자의 행동을 인식하고자 하는 '사용자 행동 인식(Human Activity Recognition)' 연구가 활발하게 진행되고 있다. 개인 사용자의 신체 일부의 단순한 움직임에 대한 인식부터, 걷거나 앉는 등의 저수준(Low-level) 행동 인식(Lukowicz et al., 2002; Lara et al., 2012), 더 나아가서는 수면을 취하거나 식사를 하는 등 보다 복잡성이 높은 고수준(High-level) 행동에 대한 인식(Vinh et al., 2011)에 이르기까지 그 영역이 확장되고 있다.

그러나 지금까지 이루어진 상당수의 사용자 행동 인식 연구가 스마트폰 사용자 개인의 행동에 초점이 맞춰져 있었으며, 사용자가 타인과 동행하고 있는지, 혹은 다른 사용자와 대화하는 중에 있는지 등 타인과의 상호 작용 행동 인식에 대한 연구는 상대적으로 미진하였다. 관련 연구에 따르면 현대인이 주변 사람들과 함께하면서 그들과 상호작용을 하는 등 사회 활동에 할애하는 시간은 전체 24시간 중 평균 8시간 이상에 달하는 것으로 조사되었다(Lee et al., 2012). 이는 스마트폰 사용자의 행동 인식 연구에 있어 주위 사람들과의 상호작용에 대한 연구가 가치를 지닐 수 있음을 방증한다.

스마트폰에는 여러 종류의 센서가 내장되어 있어서 사용자의 생활 속에서 발생하는 다종 데이터를 실시간으로 수집하는 것이 가능하며, 이는 사용자 행동 인식의 기반이 된다. 스마트폰에는 가속도계(Accelerometer), 자기장(Magnetic Field), 자이로스코프(Gyroscope) 등의 기본적인 물리 센서를 비롯하여, 오디오, 주변 블루투스 기기, 와이파이, GPS 등의 추가적인 센서가 탑재되어 있다. 기기에서 지정된 수집 빈도에 따라 이들 센서 데이터가 주기적으로 수집되며, 사용자의 권한 부여 여부에 따라 다른 응용 애플리케이션에서 해당 데이터에 접근하고 이를 필요에 맞게 가공하여 사용할 수 있도록 구성되어 있다.

스마트폰 사용자의 상호작용 행동 인식에 대한 연구는 스마트폰에서 수집하는 오디오, 블루투스, 와이파이 등의 센서에서 수집한 데이터에 의존하였다(Lu et al., 2011; Tarzia et al., 2011; Xu et al., 2013; Liu et al., 2014; Enrique et al., 2014). 이들 센서에서 수집한 데이터의 경우, 스마트폰 기기 주변의 컨텍스트를 명시적으로 내포하고 있으나, 그만큼 스마트폰 사용자의 사생활 침해 가능성이 높다. 또한 센서의 특성상 물리 센서에 비해 수집 빈도가 낮은 편이기 때문에, 충분한 양의 데이터를 확보하는 데 상대적으로 긴 시간이 소요된다. 반면에 가속도계, 자기장, 자이로스코프 등의 물리 센서의 경우, 수집 빈도가 다른 센서에 비해 월등하게 높기 때문에 단시간 내에 풍부한 데이터를 수집할 수 있으며, 기본적인 물리량만을 수집하기 때문에 사생활 침해 가능성이 상대적으로 낮다.

본 연구에서는 오디오, 블루투스, 와이파이 센서 데이터가 가지고 있는 문제점을 해결하면서, 기존 사용자 상호작용 행동 인식 연구들에서 보인 한계를 극복하기 위해, 스마트폰 상의 다종 물리 센서에서 수집한 데이터만을 사용하는 딥러닝 모델에 기반을 둔 사용자의 동행 상태 인식 모델을 제시한다. 이때의 동행 상태란 사용자 상호작용 행동의 일부를 재정의한 것으로, 사용자가 지인과 가까운 거리 안에서 동행하고 있는지와, 주변 지인과 적극적으로 대화하고 있는지를 포함한 것으로 정의한다.

본 논문의 구성은 다음과 같다. 2장에서는 사용자 동행 상태 인식 관련 연구를 소개하고, 3장에서는 스마트폰 상의 다종 물리 센서 데이터를 이용한 사용자 동행 상태 인식 모델을 제시한다. 4장에서는 실제 수집한 데이터를 이용한 실험 및 실험 결과에 대해 기술하고, 마지막 5장에서는 결론 및 향후 연구에 대해 다룬다.

## 2. 관련 연구

스마트폰 사용자가 타인과 동행하고 있는지를 인식하고자, 사용자의 대화 상황 인식에 대한 연구가 진행되어 왔다. 이들 중 상당수의 연구에서 사용자의 대화 상황 인식을 위해 오디오 데이터에 의존하는 경향을 보였다(Lu et al., 2011; Tarzia et al., 2011; Xu et al., 2013). 그러나 이러한 연구들의 경우, 사용자가 대화를 하지 않으면 타인과 동행하고 있는 상황에 대한 인식이 어려울 것이라고 짐작할 수 있다. 한편 대화 상황에 대한 인식을 거치지 않고 타인과의 동행 상황을 인식하고자 시도한 연구들도 진행되어 왔다. 이들 중 대부분은 동행 상황 인식을 위해 블루투스 혹은 와이파이 센서 데이터를 사용하였다(Liu et al., 2014; Enrique et al., 2014). 그러나 블루투스 데이터나 와이파이 데이터를 사용하여 한 사용자의 동행 상황을 인식하고자 하는 경우, 해당 사용자 주변의 모든 스마트폰에서 수집되는 블루투스 데이터나 와이파이 데이터를 동시에 사용해야 한다는 어려움이 있다. 뿐만 아니라, 사용자만 스마트폰을 소유하고 있고 그의 동행 상대자가 스마트폰을 소유하고 있지 않은 경우, 블루투스 데이터나 와이파이 데이터를 사용하여 동행 상황을 인식하는 것 자체가 불가능하다.

가속도, 자기장, 자이로스코프 등의 물리 센서 데이터의 경우 시그널 데이터로서의 특징을 지니기 때문에 모델을 학습하는 데 사용하기에 앞서 이들 데이터를 유효한 형태로 변환하고 효과적인 요인을 추출하는 과정이 반드시 선행되어야 한다. 이에 따라 물리 센서 데이터를 사용하는 사용자 행동 인식 연구들에서는 이러한 과정을 비중 있게 다뤘다(Davide et al., 2010).

이미지 분류와 음성 인식 등의 분야에서 우수한 요인 추출 능력으로 각광받고 있는 딥러닝 모델(Frank et al., 2011; Alex et al., 2012)에 대한 관심이 행동 인식 연구 분야에서도 점차 커지고 있다. 물리 센서 데이터와 같은 시그널 데이터에 딥러닝 모델을 적용함으로써, 효과적으로 잠재 요인을 추출하고 이를 통해 사용자 행동 인식 연구에서 월등한 성능을 거둔 사례가 등장하고 있다(Jiang et al., 2015; Ronao et al., 2015; Chen et al., 2016; Ordóñez et al., 2016).

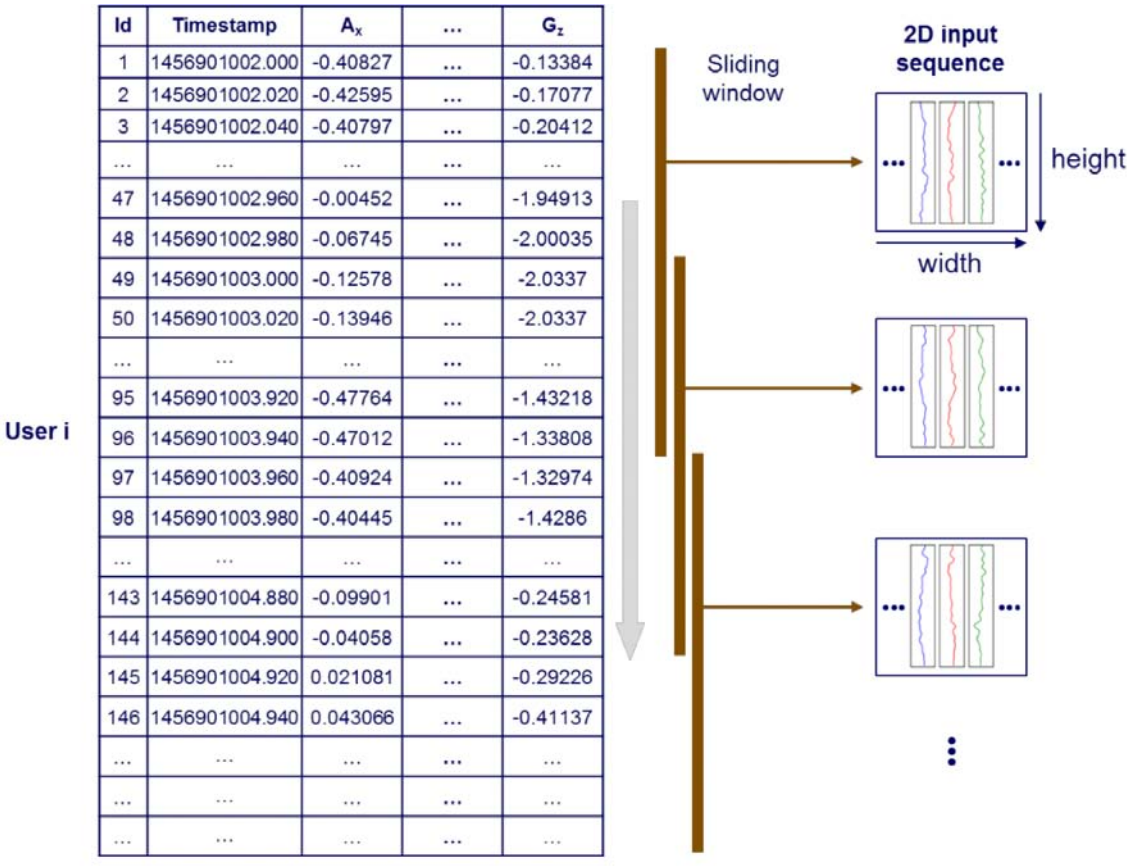
본 연구에서는 스마트폰의 가속도, 자기장, 자이로스코프 등의 다종 물리 센서에서 수집한 데이터를 입력으로 받아, 사용자의 동행 여부 및 대화 여부를 분류하는 딥러닝 모델 기반의 동행 상태 인식 모델을 제안한다.

## 3. 제안 방법론

### 3.1 데이터 전처리

스마트폰의 가속도, 자기장, 자이로스코프 등의 물리 센서에서 시간에 따라 독립적으로 수집한 데이터를 서로 비교해 보면 이들 간의 수집 타임 스탬프(Time Stamp)가 서로 완전히 일치하지 않고 미세한 차이가 나타난다. 이러한 차이

를 상쇄시키면서 이들을 동일한 시간 간격 타임스탬프들로 구성된 단일 시간 축으로 종합하기 위해, 근접 보간법(Nearest Interpolation)을 응용하여(Stisen et al., 2015) 각 센서 데이터의 수집 시점 값을 비교하고 이를 기반으로 전체 인스턴스들을 하나의 인스턴스로 합쳤다. 그리고 사용자들의 기기 간 수집 데이터 값의 차이를 상쇄시키기 위하여 각 센서 데이터 x, y, z 축 값별로 전체 사용자에 대한 정규화(Normalization)를 수행하였다. 마지막으로 결측값을 제거하고, <Figure 1>과 같이 슬라이딩 윈도우(Sliding Window) 방식에 따라 일정한 길이의 시퀀스 형태의 데이터로 재구성하였다. 원도우의 길이를 원본 데이터 인스턴스 96개에 해당하는 1920ms로 고정한 상태로 해당 작업을 진행하였으며, 50% 오버래핑(Overlapping)을 적용하여 각 시퀀스 데이터 인스턴스를 생성하였다. 이렇게 생성된 시퀀스 데이터는 이미지 데이터와 같이 2차원 구조를 지니게 되어, 컨볼루션 신경망(CNN, Convolutional Neural Network)(LeCun et al., 1995)의 입력층에 입력될 수 있는 형태가 갖춰진다.

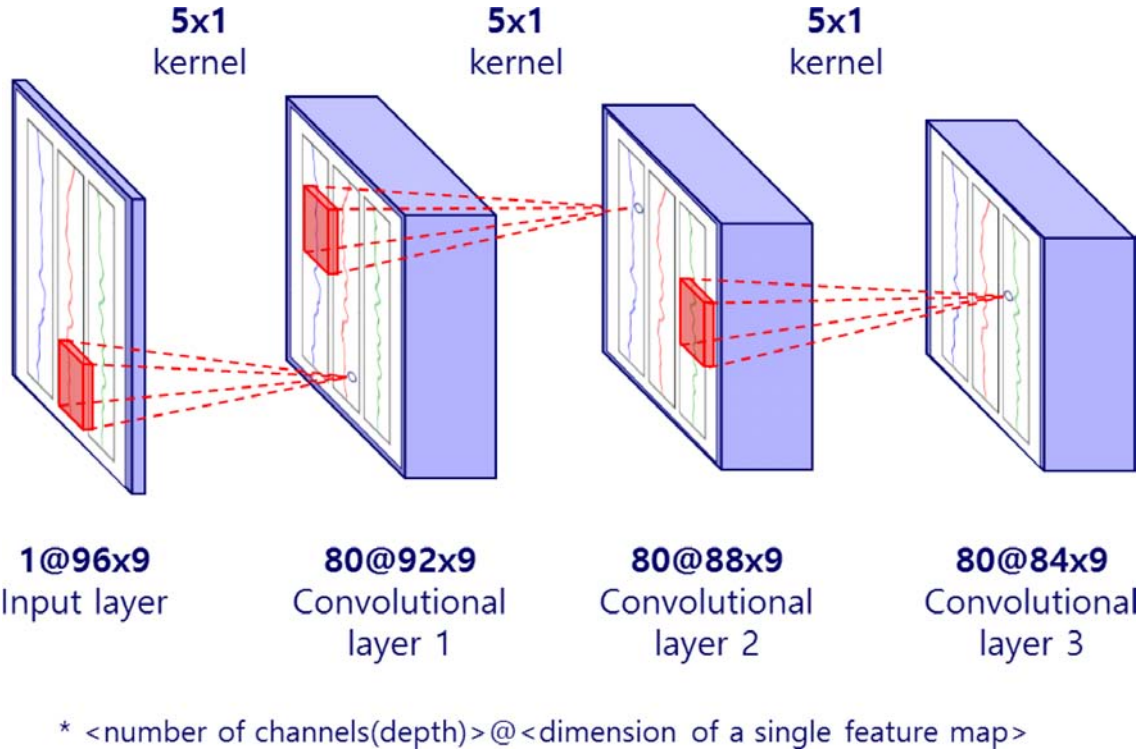



〈Figure 1〉 Example of Converting Physical Sensor Data

### 3.2 컨볼루션 신경망

전처리가 끝난 데이터는 컨볼루션 신경망의 입력층으로 보내진다. <Figure 2>는 컨볼루션 신경망의 전체 구조를 나타낸다. 컨볼루션 신경망은 세 개의 컨볼루션 층(Convolutional Layer)으로 구성되어 있으며 일반적인 컨볼루션 신경망 구조(Alex et al., 2012)와는 달리 통합층(Pooling Layer)을 두고 있지 않다. 이는 컨볼루션 연산이 이루어진 시퀀스 데이터의 시간적 위치 정보를 최대한 유지하여, 뒤에 이어질 장단기 기억 순환 신경망에서 시퀀스 데이터의 순차적 연관 관계를 효과적으로 학습할 수 있도록 하기 위함이다.

컨볼루션 층에서의 커널은 2차원 입력 데이터의 수직 방향에 해당하는 시간 축 방향에 대해서만 컨볼루션 연산을 수행하며 커널의 크기는 5로 고정하였다. 컨볼루션 연산을 거친 결과에 선형 정류 유닛(Rectified Linear Unit) 함수를 적용하여 요인 지도(Feature Map)의 값을 계산한다. 입력층에서는 2차원 입력 데이터가 단일 요인 지도로써 입력되나, 매 컨볼루션 층에서 생성되는 요인 지도의 개수는 80개로 동일하게 유지된다. 결과적으로 세 번째 컨볼루션 층에서 출력한 각 요인 지도는, 반복된 컨볼루션 연산으로 인해 2차원 입력 데이터와 비교하면 수직 방향 길이가 일부 축소된 형태를 지니게 된다.




〈Figure 2〉 CNN Structure of Proposed Model

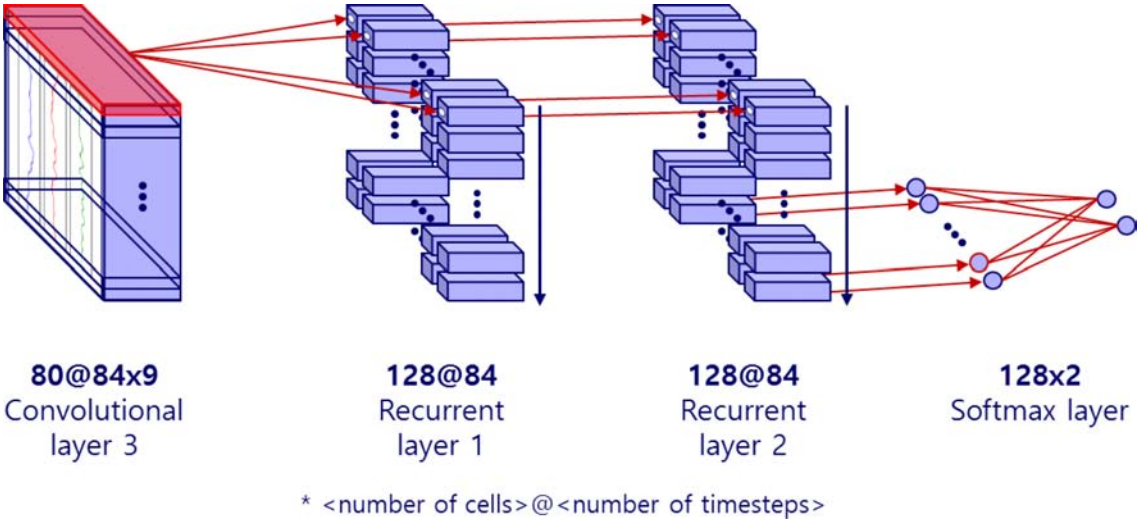
### 3.3 장단기 순환 신경망 및 최종 분류기

아래의 <Figure 3>는 장단기 순환 신경망과 최종 분류기의 구조를 나타낸다. 장단기 기억 순환 신경망(LSTM, Long Short-Term Memory Recurrent Network)(Hochreiter et al., 1997)은 두 개의 순환 신경망 층으로 구성되어 있으며, 각 순환 신경망 층에는 복수 개의 장단기 기억 유닛이 존재한다. 컨볼루션 신경망의 세 번째 컨볼루션 층의 출력값은 시간 축을 기준으로 분절되어 장단기 기억 순환 신경망 내 첫 번째 순환 신경망 층에 반복적으로 입력된다. 매 입력 시, 각 요인 값들은 모든 장단기 기억 유닛들에 각각 입력된다. 각 장단기 기억 유닛에서의 출력값 계산 시에는 쌍곡탄젠트(Hyperbolic Tangent) 함수가 적용된다. 장단기 기억 순환 신경망의 두 번째 순환 신경망 층의 각 장단기 기억 유닛은, 첫 번째 순환 신경망 층의 각 장단기 기억 유닛의 시간에 따른 출력값을 입력값으로 받아들인 뒤, 첫 번째 순환 신경망 층에서와 동일한 방식으로 시간에 따른 출력값을 계산한다.

장단기 기억 순환 신경망 내 두 번째 순환 신경망 층의 각 장단기 기억 유닛에서 맨 마지막으로 출력된 값들은, 한 층의 신경망을 추가로 거친 뒤 소프트맥스 분류기를 통해 최종 분류 결과를 도출한다.

### 3.4 모델 학습

동행 상태 인식 모델에서 소프트맥스 분류기에 의해 분류된 결과의 오류(Error)는 교차 엔트로피(Cross Entropy) 손실 함수를 사용하여 계산한다. 그리고 계산된 오류를 최소화하는 방향으로 모델의 전체 가중치를 조정하는데, 이때 adaptive moment estimation(ADAM) 알고리즘(Kingma et al., 2014)을 적용하며, 미니 배치의 크기는 128로 설정한다. 초기 학습률(Learning Rate)은 0.001로 설정하며 1회 시기(Epoch)의 학




〈Figure 3〉 LSTM Recurrent Network and Last Classifier of Proposed Model

습을 완수할 때마다 학습률이 지수함수적으로 붕괴(Exponential Decay) 하도록 하여 매 시기마다 0.99의 비율로 감소시킨다. 이를 통해 학습 과정에서 시간이 지날수록 하나의 최적해로의 수렴 가능성을 점차적으로 높인다. 모델에서 사용된 모든 가중치는 평균이 0이고 표준편차가 0.1인 정규분포 상에서 임의로 추출하여 초기화한다. 학습 데이터에 대한 과적합을 방지하기 위하여 동행 상태 인식 모델을 학습할 시, 장단기 기억 순환망 층의 입력값에 드롭아웃(Dropout) 확률 0.3의 드롭아웃 기법(Srivastava et al., 2014)을 적용한다.

## 4. 실험 및 결과

### 4.1 실험 데이터

본 연구에서 제안한 스마트폰 사용자의 동행 상태 인식 모델의 성능을 평가하기 위하여, 실험용 데이터 수집 애플리케이션 SCDC(Smart Campus Data Collection)을 개발하였다. <Figure 4>는 SCDC의 주요 화면을 나타낸다. SCDC는 실험을 수행중인 경우에 한해서만 스마트폰 상의 물리 센서를 구동할 수 있도록 제어하는 기능을 지원한다. 그리고 피실험자로 하여금 동행, 대화와 같이 현재 수행하고 있는 행동을 버튼 클릭 방식으로 손쉽게 레이블링할 수 있도록 구성되어 있다. 이 애플리케이션을 피실험자들에게 배포하여 스마트폰 상의 가속도, 자기장, 자이로스코프 센서 데이터를 50Hz의 주기로 수집하였다.

총 18명의 피실험자들을 대상으로 스마트폰 데이터 수집 실험을 수행하였다. 이들은 모두 대학생들이며 자연스러운 대화 상황에서의 데이터


〈Figure 4〉 the main UI of SCDC application

를 수집하기 위하여 지인들로 구성되었다. 피실험자들은 3주간의 실험 기간에 일 인당 평균적으로 12.5회 실험에 참여하였다. 먼저 동행 중 및 대화 중 데이터 수집 실험의 경우 실험자의 지도 및 감독 하에 이루어졌다. 대학교 캠퍼스 내에 세 가지 경로를 지정한 뒤, 두 명 이상의 피실험자로 하여금 하나의 그룹을 이루도록 한 뒤, 이들이 SCDC를 실행한 상태로 서로 대화를 하거나 혹은 대화를 하지 않으면서 지정한 경로를 왕복하도록 하였다. 한편 비 동행 중 및 비 대화 중 데이터 수집 실험의 경우 실험자의 감독 없이, 피실험자의 자율적인 참여로 이루어졌다. 피실험자가 캠퍼스를 동행인 없이 홀로 이동하는 경우, SCDC를 실행한 상태로 이동하도록 지시하

〈Table 1〉 Summary of Collected Smartphone Multimodal Data

| Smartphone Sensor | Data Size (MB) | The Number of Data Records | | | | |
|---|---|---|---|---|---|---|
| | | Total Number of Data Records | Non-Group Interaction | Group Interaction | Non-Conversation | Conversation |
| Accelerometer | 865.63 | 7,155,446 | 468,216 | 3,515,412 | 469,942 | 3,508,674 |
| Magnetic Field | 803.61 | 6,553,904 | 399,872 | 3,042,017 | 401,704 | 3,035,757 |
| Gyroscope | 636.59 | 5,323,297 | 302,796 | 2,828,841 | 304,590 | 2,823,906 |

였다. 모든 실험 과정에서 피실험자들의 행동을 과도하게 제한하지는 않았으며, 평소와 같이 자연스럽게 행동하도록 하였다. 그리고 데이터를 수집하는 스마트폰의 위치를 특정 신체 부위에 제약하지 않았다.

전체 피실험자로부터 수집한 스마트폰 다종 데이터를 요약한 결과는 위의 <Table 1>과 같다. 동행과 비 동행, 대화와 비 대화 상태를 분류하기 위하여 가속도, 자기장 그리고 자이로스코프 데이터를 수집하였다. 수집한 데이터에 대해 전처리를 수행하여 시퀀스 데이터로 변환하고 이들을 임의의 순서로 섞은 뒤, 6:2:2의 비율로 나누어 학습 데이터, 검증 데이터, 평가 데이터를 구성하였다.

### 4.2 실험 결과

사용자 동행 상태, 특히나 동행 여부와 대화 여부의 인식을 위해 동행 데이터와 대화 데이터 각각에 대하여 제안 모델과 베이스라인 모델들을 학습하여 모델들 간의 성능을 비교하였다. 성능 비교를 위한 베이스라인 모델로는 다수결 분류기(Majority Vote Classifier) 모델을 비롯하여, 지지 벡터 기계(Support Vector Machine), 본 연구에서 제안하는 동행 상태 인식 모델의 구조를 일부 변형한 심층 순환 신경망(Deep Recurrent Neural Network) 모델, 심층 장단기 순환 신경망(Deep LSTM Recurrent Network) 모델을 사용하였다. 지지 벡터 기계의 경우, 단일 시퀀스 데이터의 평균, 표준편차, 최댓값, 최솟값, 중간값 등의 기본적인 통계량을 계산하고 이들을 요인으로 사용하였다. 심층 순환 신경망 모델과 심층 장단기 순환 신경망 모델은 본 연구에서 제안한 동행 상태 인식 모델에서 컨볼루션 신경망 부분을 제거하고, 각각 순환 신경망 유닛과 장단기 기억 유닛을 사용한 모델이다.

본 연구에서 수집한 데이터에는 동행/대화 클래스와 비 동행/비 대화 클래스 간에 수적 불균형이 존재하기 때문에, 다수 클래스로 모든 인스턴스를 단순 일괄 분류하는 다수결 분류기 모델과 비교하여 얼마나 더 우수한 성능을 보일 수 있는지가 유의미한 지표라고 할 수 있다. 한편 동행 상태 인식 모델 구조의 일부를 공유하는 베이스라인 모델들의 경우, 미니 배치의 크기, 초기 학습률 등의 파라미터들을 동행 상태 인식 모델과 동일하게 설정하였다. 동행 상태 분류를 위한 베이스라인 모델들과 본 논문에서 제안하는 컨볼루션 신경망과 장단기 기억 순환 신경망이 결합된 모델의 성능 비교 실험 결과가 아래 <Table 2>와 같다.

〈Table 2〉 Error Rate and F1 Score Comparison between Proposed Model and Baseline Model

| Model | Group Interaction | | Conversation | |
|---|---|---|---|---|
| | Error Rate (%) | F1 Score | Error Rate (%) | F1 Score |
| Majority Vote Classifier | 10.606 | 0.844 | 10.789 | 0.841 |
| Deep Recurrent Neural Network | 7.021 | 0.919 | 10.070 | 0.861 |
| Support Vector Machine | 6.218 | 0.927 | 6.256 | 0.926 |
| Deep LSTM Recurrent Network | 2.536 | 0.974 | 2.539 | 0.974 |
| **Ours** | **1.263** | **0.987** | **1.168** | **0.988** |

실험 결과, 데이터를 기반으로 학습된 모델들이 모두 다수결 분류 모델보다 인식 성능이 우수하였으며, 지지 벡터 기계 모델이 심층 순환 신경망 모델보다 성능이 우수하였다. 그리고 심층 장단기 순환 신경망 모델의 성능이 지지 벡터 기계 모델보다 더 우수하였는데, 이는 장단기 기억 유닛이 동행 상태 인식에 필요한 물리 센서 시퀀스 데이터의 시간에 따른 순차적 연관 관계뿐만 아니라, 장기 의존성도 학습하여 효과적으로 활용하였기 때문인 것으로 분석할 수 있다.

본 연구에서 제안한 동행 상태 인식 모델이 가장 높은 성능을 보였는데, 심층 장단기 순환 신경망 모델과 비교할 때 이는 물리 센서 시퀀스 데이터에 대해 컨볼루션 연산을 적용한 데 기인한다. 순환 신경망에서는 포착할 수 없는 국부 영역에서의 기하학적 의존성을 복수 개의 커널이 다양한 각도에서 포착하고 이를 요인 지도의 형태로 확장하여 출력함으로써, 이를 입력값으로 받는 장단기 기억 순환 신경망에서 보다 풍부한 정보를 반영할 수 있었던 것으로 분석할 수 있다.

딥러닝 모델의 경우, 우수한 성능을 보이더라도 모델의 복잡한 비선형 구조로 인하여 데이터 차원에서 그 요인을 파악하는 것은 어려운 일이다(Samek et al., 2017). 하지만 그 요인들을 유추하자면 다음과 같은 요인들이 있을 수 있다. 스마트폰 사용자가 타인과 동행 중인 경우, 사용자의 걸음 패턴이나 신체의 회전 정도가 비 동행 중인 경우와 미세한 차이가 발생하여 물리 센서 값들이 변화했을 가능성이 있다. 예를 들어, 사용자가 타인과 동행 중인 경우 동행인의 걷는 상태를 확인하기 위해 신체를 동행인 방향으로 자주 회전시켰을 것이며 이는 자이로스코프와 자기장 센서의 값을 변화시켰을 것이다. 그리고 사용자가 타인과 걷는 속도를 맞추기 위해 자신의 걸음을 의식적으로 조절하여 가속도계의 값이 크게 변화했을 것이라 추측할 수 있다. 이와 유사하게, 사용자가 타인과 대화 중인 경우에도 비 대화 중인 경우에 비해 팔/다리 등의 신체 부위의 움직임에 차이가 발생하였을 것이며, 말을 하는 과정에서 신체에서 미세한 진동이 발생하여 물리 센서 값이 변화했을 것이다.

본 연구에서 제안한 동행 상태 인식 모델은 물리 센서 시퀀스 데이터에서 감지되는 이러한 미

세한 변화를 포착하고 이를 요인으로 추출함으로써, 동행 중 및 대화 중의 오분류율이 각각 1.263%, 1.168%이고 F1 점수가 0.987, 0.988인 높은 성능을 거둘 수 있었던 것으로 분석할 수 있다.

## 5. 결론 및 향후 연구

본 연구에서는 스마트폰 상의 가속도, 자기장, 자이로스코프 등의 다종 물리 센서에서 수집한 데이터만을 바탕으로, 딥러닝 기반의 동행 상태 인식 모델을 사용하여 사용자의 동행 여부 및 대화 여부를 인식하였다. 가속도, 자기장, 자이로스코프 등의 물리 센서는 기존 연구에서 사용했던 오디오, 블루투스, 와이파이 등의 센서 데이터에 비해 사용자의 사생활 침해 가능성이 낮으며, 단시간 내에 많은 양의 데이터를 수집할 수 있다는 장점이 있다.

동행 상태 분류 모델을 학습하기 위해, 먼저 스마트폰 상의 각 물리 센서로부터 시간에 따라 각각 독립적으로 수집한 데이터를 타임 스탬프를 맞추어 하나의 입력 데이터로 종합하고, 결측값을 제거하며 정규화를 수행한 뒤 시퀀스 데이터로 변환하는 데이터 전처리 과정을 거쳤다. 이어지는 컨볼루션 신경망에서는 입력된 시퀀스 데이터의 국부 의존성을 학습하여, 이를 요인 지도의 형태로 추출하였다. 장단기 기억 순환 신경망은 요인 지도를 입력받아 시간에 따른 순차적 연관 관계를 학습하며, 이 과정에서 시퀀스 데이터의 장기 의존성을 반영하여 최종적인 요인을 추출하였다. 마지막 단계에서의 소프트맥스 분류기는 해당 요인을 사용하여 동행/대화 여부에 대한 최종 분류를 수행하였다. 이와 같이 동행 상태 분류 모델은 컨볼루션 신경망과 장단기 기억 순환 신경망의 장점을 융합하여, 단순한 물리 센서 데이터만을 사용하고도 동행과 대화라는 고수준 행동 인식에 있어서 각각 0.987, 0.988의 높은 정확도를 기록할 수 있었다.

동행 상태 분류 모델은 전통적인 분류 모델에 비해 가중치의 수가 많아 복잡성이 상대적으로 높으며, 이 때문에 학습에 필요한 데이터의 양도 더 많다. 그럼에도 불구하고 가속도, 자기장, 자이로스코프 등의 물리 센서는 데이터 수집을 위해 사용자의 권한 승인을 요구하지 않으며, 수집 과정에서 요구되는 계산량 및 에너지 소모량도 상대적으로 적다. 따라서 모델 학습을 위한 데이터 수집에 있어 다른 센서 데이터에 비해 어려움이 적다고 할 수 있다. 최종 학습된 동행 상태 분류 모델을 다른 스마트폰 애플리케이션에 이식하여 실제 서비스에 활용하고자 하는 경우에도, 위에서 언급한 물리 센서의 장점은 그대로 유효하다. 애플리케이션 사용자의 권한 승인 없이, 스마트폰 운영체제에 부담을 주지 않는 수준에서 물리 센서 데이터를 수시로 짧은 시간 동안 수집하는 것이 가능하며, 이를 통해 사용자가 어느 시점에 타인과 상호작용을 하고 있는지 인식함으로써 이에 맞는 시의적절한 서비스를 제공할 수 있다.

본 연구에서는 스마트폰 상에서 수집한 물리 센서 데이터를 동행 상태 분류 모델에 입력할 수 있는 시퀀스 데이터로 변환하는 데이터 전처리 작업을 선행하였다. 그러나 데이터 전처리 작업 중 서로 다른 종류의 센서 데이터 간의 타임 스탬프를 일치시키는 현재의 방법에서는, 비록 ms 단위의 짧은 시간 간격일지라도 시간적인 오차가 발생하는 것이 불가피하다. 이로 인해, 데이터 전처리 결과 생성된 시퀀스 데이터는 스마트

폰 상에서 감지된 가속도, 자기장, 자이로스코프의 실제 시간에 따른 패턴을 어느 정도 왜곡시켰을 가능성이 존재한다. 향후 연구에서는 이러한 왜곡을 최소화할 수 있는 보다 엄밀한 다종 센서 데이터 종합 방법을 연구할 예정이다. 또한, 학습 데이터에 맞춰 학습된 모델을 기존과 다른 분포를 따르는 평가 데이터에 맞게 전이할 수 있도록 하는 전이 학습(Transfer Learning) 방법을 추가적으로 연구할 예정이다. 이를 통해, 모델 학습 단계에서 고려하지 못한 데이터 상의 변화에 대해서도 강건한 인식 성능을 발휘할 수 있는 모델을 얻을 수 있을 것으로 기대한다.

## 참고문헌(References)


Alex, K., I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," *Proceedings of neural information processing systems*, (2012), 1097~1105.

Chen, Y., K. Zhong, J. Zhang, Q. Sun, and X. Zhao, "LSTM Networks for Mobile Human Activity Recognition," *Proceedings of International Conference on Artificial Intelligence: Technologies and Applications*, (2016), 50~53.

Davide, F, P. C. Diniz, D. R. Ferreira, and J. M. Cardoso, "Preprocessing techniques for context recognition from accelerometer data," *Personal and Ubiquitous Computing*, Vol. 14, No. 7(2010), 645~662.

Enrique, G., V. Osmani, A. Maxhuni, and O. Mayora, "Detecting Walking in Synchrony Through Smartphone Accelerometer and Wi-Fi Traces," *Proceedings of AmI 2014: Ambient Intelligence*, (2014), 33~46.

Frank, S., G. Li, X. Chen, and D. Yu, "Feature engineering in context-dependent deep neural networks for conversational speech transcription," *Proceedings of IEEE Workshop Automatic Speech Recognition and Understanding*, (2011), 24~29.

Hochreiter, S., and J. Schmidhuber, "Long short-term memory," *Neural computation*, Vol. 9, No. 8(1997), 1735~1780.

Jiang, W., and Z. Yin, "Human activity recognition using wearable sensors by deep convolutional neural networks," *Proceedings of the 23rd ACM international conference on Multimedia*, (2015), 1307~1310.

Kingma, D. P., and J. L. Ba, "Adam: A method for stochastic optimization," *arXiv preprint arXiv: 1412.6980*(2014).

Lara, O. D., and M. A. Labrador, "A mobile platform for real-time human activity recognition," *Proceedings of Consumer Communications and Networking Conference*, (2012), 667~671.

LeCun, Y., and Y. Bengio, "Convolutional networks for images, speech, and time series," *The handbook of brain theory and neural networks*, Vol. 3361, No. 10(1995), 255~258.

Lee, Y., Y. Ju, C. Min, S. Kang, I. Hwang, and J. Song, "Comon: Cooperative ambience monitoring platform with continuity and benefit awareness," *Proceedings of the 10th international conference on Mobile systems, applications, and services*, (2012), 43~56.

Liu, S., Y. Jiang, and A. Striegel, "Face-to-face proximity estimationusing bluetooth on smartphones," *IEEE Transactions on Mobile Computing*, Vol. 13, No. 4(2014), 811~823.

Lu, Hong, A. B. Brush, B. Priyantha, A. K. Karlson, and J. Liu, "Speakersense: Energy efficient unobtrusive speaker identification on mobile phones," *Proceedings of Pervasive Computing*, (2011), 188~205.

Lukowicz, P., H. Junker, M. St¨ager, T. von B¨uren, and G. Tr¨oster, "WearNET: A distributed multi-sensor system for context aware wearables," *Proceedings of UbiComp 2002: Ubiquitous Computing*, (2002), 361~370.

Ordóñez, F. J., and D. Roggen, "Deep convolutional and lstm recurrent neural networks for multimodal wearable activity recognition," *Sensors*, Vol. 16, No. 1(2016), 115.

Ronao, C. A., and S. Cho, "Deep convolutional neural networks for human activity recognition with smartphone sensors," *Proceedings of International Conference on Neural Information Processing*, (2015), 46~53.

Samek, W., T. Wiegand, and K. Müller, "Explainable artificial intelligence: Understanding, visualizing and interpreting deep learning models," *arXiv preprint arXiv: 1708.08296*(2017).

Srivastava, N., G. Hinton, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov, "Dropout: a simple way to prevent neural networks from overfitting," *Journal of machine learning research*, Vol. 15, No. 1(2014), 1929~1957.

Stisen, A., H. Blunck, S. Bhattacharya, T. S. Prentow, M. B. Kjærgaard, A. Dey, T. Sonne, and M. M. Jensen, "Smart devices are different: Assessing and miti-gatingmobile sensing heterogeneities for activity recognition," *Proceedings of the 13th ACM Conference on Embedded Networked Sensor Systems*, (2015), 127~140.

Tarzia, S. P., P. A. Dinda, R. P. Dick, and G. Memik, "Indoor localization without infrastructure using the acoustic background spectrum," *Proceedings of the 9th international conference on Mobile systems, applications, and services*, (2011), 155~168.

Vinh, L. T., S. Lee, H. X. Le, H. Q. Ngo, H. I. Kim, M. Han, and Y. Lee, "Semi-Markov conditional random fields for accelerometer-based activity recognition," *Applied Intelligence*, Vol. 35, No. 2(2011), 226~241.

Xu, C., S. Li, G. Liu, Y. Zhang, E. Miluzzo, Y. Chen, J. Li, and B. Firner, "Crowd++: unsupervised speaker count with smartphones," *Proceedings of the 2013 ACM international joint conference on Pervasive and ubiquitous computing*, (2013), 43~52.

Abstract

# A Deep Learning Based Approach to Recognizing Accompanying Status of Smartphone Users Using Multimodal Data

Kilho Kim* · Sangwoo Choi** · Moon-jung Chae***
· Heewoong Park*** · Jaehong Lee**** · Jonghun Park***

As smartphones are getting widely used, human activity recognition (HAR) tasks for recognizing personal activities of smartphone users with multimodal data have been actively studied recently. The research area is expanding from the recognition of the simple body movement of an individual user to the recognition of low-level behavior and high-level behavior. However, HAR tasks for recognizing interaction behavior with other people, such as whether the user is accompanying or communicating with someone else, have gotten less attention so far. And previous research for recognizing interaction behavior has usually depended on audio, Bluetooth, and Wi-Fi sensors, which are vulnerable to privacy issues and require much time to collect enough data. Whereas physical sensors including accelerometer, magnetic field and gyroscope sensors are less vulnerable to privacy issues and can collect a large amount of data within a short time. In this paper, a method for detecting accompanying status based on deep learning model by only using multimodal physical sensor data, such as an accelerometer, magnetic field and gyroscope, was proposed. The accompanying status was defined as a redefinition of a part of the user interaction behavior, including whether the user is accompanying with an acquaintance at a close distance and the user is actively communicating with the acquaintance. A framework based on convolutional neural networks (CNN) and long short-term memory (LSTM) recurrent networks for classifying accompanying and conversation was proposed.

First, a data preprocessing method which consists of time synchronization of multimodal data from

* SUALAB
** Corresponding Author: Sangwoo Choi
Recommendations, Coupang
19F, 570, Songpadae-ro, Songpa-gu, Seoul, Korea
Tel: +82-2-6150-4688, E-mail: sangwoochoi@coupang.com
*** Department of Industrial Engineering and Institute for Industrial Systems Innovation, Seoul National University
**** kakaomobility datalab

different physical sensors, data normalization and sequence data generation was introduced. We applied the nearest interpolation to synchronize the time of collected data from different sensors. Normalization was performed for each x, y, z axis value of the sensor data, and the sequence data was generated according to the sliding window method. Then, the sequence data became the input for CNN, where feature maps representing local dependencies of the original sequence are extracted. The CNN consisted of 3 convolutional layers and did not have a pooling layer to maintain the temporal information of the sequence data. Next, LSTM recurrent networks received the feature maps, learned long-term dependencies from them and extracted features. The LSTM recurrent networks consisted of two layers, each with 128 cells. Finally, the extracted features were used for classification by softmax classifier. The loss function of the model was cross entropy function and the weights of the model were randomly initialized on a normal distribution with an average of 0 and a standard deviation of 0.1. The model was trained using adaptive moment estimation (ADAM) optimization algorithm and the mini batch size was set to 128. We applied dropout to input values of the LSTM recurrent networks to prevent overfitting. The initial learning rate was set to 0.001, and it decreased exponentially by 0.99 at the end of each epoch training.

An Android smartphone application was developed and released to collect data. We collected smartphone data for a total of 18 subjects. Using the data, the model classified accompanying and conversation by 98.74% and 98.83% accuracy each. Both the F1 score and accuracy of the model were higher than the F1 score and accuracy of the majority vote classifier, support vector machine, and deep recurrent neural network. In the future research, we will focus on more rigorous multimodal sensor data synchronization methods that minimize the time stamp differences. In addition, we will further study transfer learning method that enables transfer of trained models tailored to the training data to the evaluation data that follows a different distribution. It is expected that a model capable of exhibiting robust recognition performance against changes in data that is not considered in the model learning stage will be obtained.

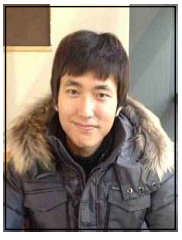
**Key Words** : human activity recognition, group interaction, smartphone multimodal sensors, convolutional neural network, long short-term memory recurrent network

Received : June 25, 2018 Revised : February 12, 2019 Accepted : March 11, 2019
Publication Type : Regular Paper Corresponding Author : Sangwoo Choi

# 저 자 소 개


**김 길 호**
서울대학교 산업공학과에서 학사 및 석사 학위를 취득하였으며, 현재 (주)수아랩의 책임연구원으로 재직 중이다. 주요 연구 분야는 automatic visual inspection, deep learning 등이다.

**최 상 우**
서울대학교 산업공학과에서 학사 및 석사 학위를 취득하였으며, 현재 (주)쿠팡 추천팀의 소프트웨어 엔지니어로 재직 중이다. 주요 연구 관심 분야는 기계학습, 추천시스템 등이다.



**채 문 정**
현재 서울대학교 산업공학과 박사과정에 재학중이다. 주요 연구 관심 분야는 음성 인식 및 합성, 딥러닝 등이다.

**박 희 웅**
현재 서울대학교 산업공학과 박사과정에 재학중이다. 주요 연구 관심 분야는 기계 독해, 자연어 이해, 사용자 프로파일링 등이다.

**이 재 홍**
한국과학기술원  전산학과에서 학사 학위를 취득하였으며, 서울대학교 산업공학과에서 석사 학위를 취득하였다. 현재 카카오모빌리티 데이터랩에서 데이터 과학자로 재직 중이며, 주요 연구 관심 분야는 사용자 프로파일링, 전이학습 등이다.

**박종헌**
서울대학교 산업공학과에서 학사와 석사를 마쳤으며, 조지아공과대학 산업공학과에서 박사 학위를 취득 후 현재 서울대학교 산업공학과 교수로 재직중이다. 주요 연구 관심 분야는 운영 애널리틱스와 사용자 모델링 등이 있다.