

# 스마트폰 센서를 이용하여 행동을 인식하기 위한 계층적인 심층 신뢰 신경망

이 현 진<sup>†</sup>

## Hierarchical Deep Belief Network for Activity Recognition Using Smartphone Sensor

Hyunjin Lee<sup>†</sup>

### ABSTRACT

Human activity recognition has been studied using various sensors and algorithms. Human activity recognition can be divided into sensor based and vision based on the method. In this paper, we proposed an activity recognition system using acceleration sensor and gyroscope sensor in smartphone among sensor based methods. We used Deep Belief Network (DBN), which is one of the most popular deep learning methods, to improve an accuracy of human activity recognition. DBN uses the entire input set as a common input. However, because of the characteristics of different time window depending on the type of human activity, the RBMs, which is a component of DBN, are configured hierarchically by combining them from different time windows. As a result of applying to real data, The proposed human activity recognition system showed stable precision.

**Key words:** Acceleration Sensor, Gyroscope Sensor, Human Activity Recognition, Deep Belief Network

### 1. 서 론

최근 스마트폰 및 웨어러블 디바이스에 대한 관심이 높아지면서 인간의 행동을 인식하는 제품들이 많은 관심을 받고 있다. 웨어러블 디바이스에는 스마트 글래스, 헤드 업 디스플레이(Head Up Display, HUD), 스마트 밴드, 스마트 워치, 스마트 의류, 스마트 신발 등이 있지만 사용자는 한정되어 있다. 하지만 스마트폰은 널리 보급되어 있기 때문에 스마트폰의 센서를 이용하면 다양한 센서들로 부터 정보를 편리하게 얻을 수 있다.

인간의 행동을 인식하는 방식은 크게 비전기반(Vision-based)의 방식과 센서기반의 방식으로 나뉜다. 비전 기반의 행동 인식은 영상 이미지를 그대로

사용하거나 카메라의 영상에 깊이 센서를 적용하여 인체의 3차원 정보를 인식하고, 인체의 관절 위치나 관절 각도와 같은 바디 부분 정보를 사용하여 행동 인식을 수행하는 것이다[1]. 센서기반의 방식은 인간의 신체에 여러 개의 카메라와 마커들을 부착한 광학 모션 센서를 이용한 마커 기반의 방식과 스마트폰이나 스마트워치 같은 범용 센서들을 사용하는 방식이 있다. 일반적인 웨어러블 디바이스들은 손목, 발, 눈 등 일정한 위치에 존재하기 때문에 행동을 인식하기 위한 센서 데이터가 일정한 패턴을 가질 수 있어서 인식이 용이하다. 하지만 스마트폰은 개인의 성향에 따라서 손이나 가방, 목 등 다양한 위치에 보관할 수 있기 때문에 센서 데이터의 패턴이 다양하게 나타날 수 있어 인식에 어려움이 있다. 따라서 스마트폰의

※ Corresponding Author : Hyunjin Lee, Address: (03132) #321, Jongno Biz-well, 23 Samil-daero 30-gil, Jongno-gu, Seoul, Korea, TEL : +82-2-708-7863, FAX : +82-2-708-7749, E-mail : hjlee@mail.kcu.ac

Receipt date : Jul. 16, 2017, Approval date : Jul. 25, 2017

<sup>†</sup> Division of ICT Engineering, Korea Soongsil Cyber University

센서를 이용하여 인간의 행동을 인식의 정확도를 높이기 위해 전처리를 수행하거나 분석알고리즘을 개발하는 연구가 진행되고 있다[2,3,4].

인공 신경회로망은 1940년대부터 연구가 시작되어 헤비안 학습(Hebbian Learning), 퍼셉트론(Perceptron)등을 거쳐 1980년대에 다층 신경회로망(Multi Layer Neural Networks)의 학습 알고리즘인 오류역전파(Error Back Propagation) 알고리즘이 만들어지면서 전성기를 맞이하였다. 기존에 사용하던 방법들을 적용하기 어려웠던 영상처리, 인공지능 등 다양한 분야에 적용되었지만, 서포트 벡터 머신(Support Vector Machine)과 같은 새로운 기계 학습 방법들이 발달하면서, 많은 도전을 받아 왔다. 2000년대 중반에 딥 러닝(Deep Learning)의 출현 이후 인공 신경회로망은 연구영역 뿐만 아니라 적용 영역에서도 큰 주목을 받고 있다. 딥러닝은 기존 방법으로는 인식률의 증가가 어려웠던 이미지 인식, 음성 인식, 영상 인식, 번역 등의 다양한 분야에서 높은 인식률을 보이고 있다.

본 논문에서는 스마트폰 센서의 데이터를 이용하여 시간 구간의 특성을 반영하여 딥러닝기반의 심층 신뢰 신경망을 구성하여 인간 행동을 인식하는 시스템을 제안한다. 널리 보급되어 있는 스마트폰에서 얻을 수 있는 센서 정보 중에서 가속도 센서(Acceleration Sensor)와 자이로스코프 센서(Gyroscope Sensor)의 데이터를 활용하여 심층 신뢰 신경망(Deep Belief Network)을 기반으로 한 인식기에 적용한다. 인간의 행동은 순간적으로 포착되는 스냅샷이 아니라 시간의 흐름에 따른 연속적인 행동의 흐름으로 정의된다. 따라서, 특정 시간에 대한 누적데이터를 사용하여 연속적인 행동의 흐름을 반영한다. 시간의 흐름도 짧은 시간의 간격과 긴 시간의 간격등을 반영하여 심층 신뢰 신경망을 계층적으로 구성하여 안정성을 높일 수 있도록 구성한다. 제한된 볼츠만 머신(Restricted Boltzmann Machine)을 이용한 사전 학습(pretraining)으로 지역 최소값을 피하는 네트워크의 초기값을 설정한 뒤 파인튜닝(Fine Tuning)으로 네트워크의 인식 성능을 높인다.

본 논문의 구성은 다음과 같이 진행된다. 2장에서 행동 인식에 관련된 연구들을 소개하고, 3장에서는 심층 신뢰 신경망을 이용하여 제안하는 방법을 소개한다. 4장에서는 실 데이터에 대한 실험을 통해 제안

하는 방법을 평가하고, 5장에서는 본 연구의 결과와 향후 연구 방향으로 결론을 맺는다.

## 2. 관련 연구

스마트폰의 센서 데이터를 이용하여 행동을 인식하는 연구들은 다음과 같다. Lee 등[5]은 스마트폰의 가속도 센서를 이용하여 데이터를 수집하고, 왼쪽, 오른쪽, 바깥쪽 등의 스마트폰의 제스처를 LSTM(Long Short Term Memory) 구조의 양방향 순환 신경망(Recurrent Neural Networks)을 이용하여 인식하는 시스템을 연구하였다.

Anguita 등[6]과 Stisen 등[7]은 스마트폰의 센서에 기록된 데이터를 이용하여 인간의 행동을 분류하고, 이 데이터를 연구 목적으로 공개하였다. 이 공개된 데이터들은 다양한 알고리즘들에 적용되어 행동 인식에 대한 연구가 이루어지고 있다. Anguita 등[6]은 서포트벡터머신(Support Vector Machine, SVM)으로 분류를 수행하였고, Reiss 등[8]은 C4.5 의사 결정 나무를 약한 분류기(weak classifier)로 사용하고, 이 결과를 AdaBoost(Adaptive Boosting) 알고리즘으로 결합한 알고리즘을 제안하였다. Chetty 등[9]은 kNN(k Nearest Neighbor) 기반의 IBk 분류기를 제안하였다. 또한, Ronao 등은 은닉 마르코프 모델(Hidden Markov Models) 기반으로 모바일 환경에 적합한 계층적인 구조를 갖는 행동 인식 방법[10]과 컨볼루션 신경망(Convolutional Neural Networks)을 이용한 분류 알고리즘[11]을 제안하였다. Jiang 등[12]은 컨볼루션 신경망이 이미지 인식에 특화되어 있으므로, 스마트폰의 센서 데이터를 이미지화 한 후 컨볼루션 신경망을 적용하는 방법을 제안하였다. Lee[13]는 인간행동인식을 위해서 센서데이터와 같은 시계열데이터를 위하여 여러 시구간에 대해 특징추출을 하여 다층 퍼셉트론으로 인식하는 다중 시구간 신경회로망 방식을 제안하였다.

## 3. 제안하는 방법

### 3.1 전체적인 시스템의 구성

제안하는 시스템의 전체적인 구성은 (Fig. 1)과 같다. 가속도 센서와 자이로스코프 센서의 입력에 대한 누적데이터를 받는다. 제한된 볼츠만 머신을 계층적으로 구성하는데 입력에 대해서 다양한 시간 간격을

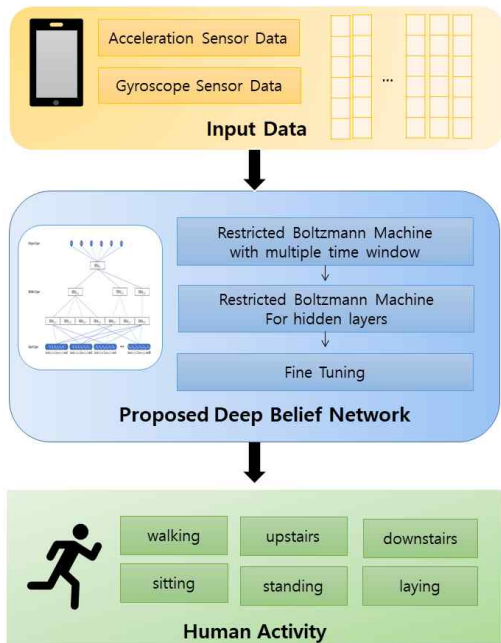


Fig. 1. Overall System Structure.

입력으로 한 제한된 볼츠만 머신을 구성하여 각각을 학습시킨다. 이들을 다시 제한된 볼츠만 머신으로 구성하여 은닉층을 구성하여 학습시키고, 최종적으로 파인튜닝 기법에 의하여 인간의 걷기(walking), 계단

오르기(upstairs), 계단 내려가기(downstairs), 앉기(sitting), 일어서기(standing), 눕기(laying)의 6가지 행동에 대한 인식결과를 생성한다.

### 3.2 센서데이터 입력과 구간데이터 생성

스마트폰에는 조도센서(Light Sensor), 지문인식 센서(Fingerprint Sensor), 이미지 센서(Image Sensor), 터치 센서(Touch Sensor), GPS 센서, 지자기 센서(Geo-magnetic Sensor), 가속도 센서(Acceleration Sensor), 자이로스코프 센서(Gyroscope Sensor) 등 다양한 센서들이 존재한다[14]. 본 논문에서 여러 가지 센서 중 모션 센서인 가속도 센서와 자이로스코프 센서를 사용하였다.

가속도 센서는 이동하는 물체의 속도 변화나 충격의 세기를 감지하고 측정하는 센서이다. 일정한 시간 내에 직선방향으로 속도 증감비를 나타내 주는 센서이다. 가속도 센서는 (Fig. 2)와 같이 3개의 축으로 구성되며 X축은 스마트폰 측면에 작용하는 가속도로 오른쪽은 +값을 왼쪽은 -값을 갖는다. Y축은 윗면과 아랫면에 작용하는 가속도로 위는 + 아래는 - 값을 갖는다. Z축은 스마트폰의 앞뒷면에 작용하는 가속도로 디스플레이 앞은 + 값을 디스플레이 뒤는 - 값을 갖는다.

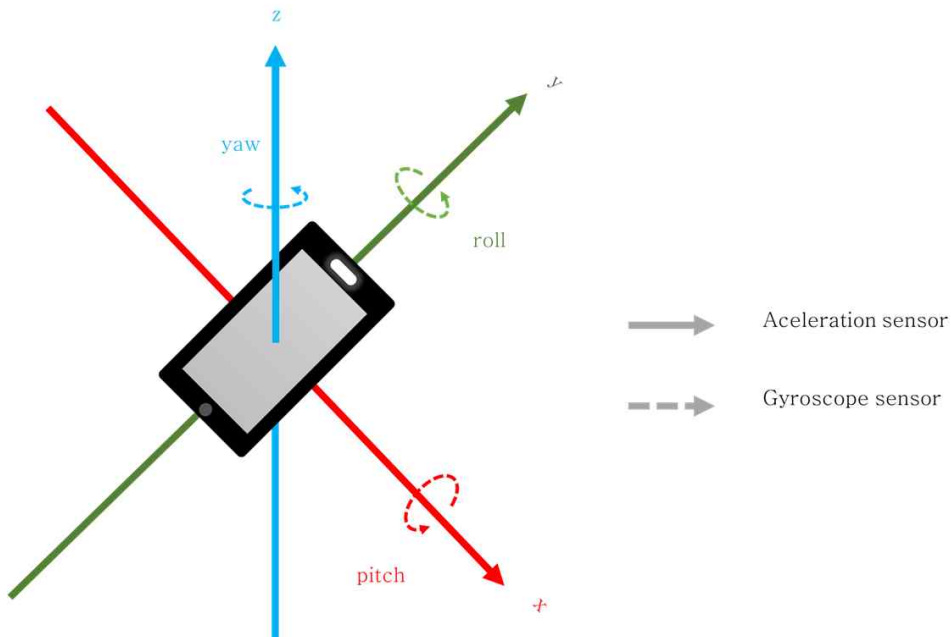


Fig. 2. Acceleration Sensor and Gyroscope Sensor.

자이로스코프센서는 물체의 회전속도인 각속도에 따른 회전방향을 알려주는 센서이다. 일정시간내에 한축을 기준으로 물체가 회전한 각도를 수치로 알려주는 센서이다. 어떤 물체가 회전 운동을 할 때에 생기는 코리올리 힘(Coriolis Force)을 전기적 신호로 변환하여 계산한다. (Fig. 2)와 같이 X축 주위의 회전을 피치(pitch)이라고 하고, Y축 주위의 회전을 롤(roll)이라고 하고, Z축 주위의 회전을 요(yaw)라 부른다.

3.3 심층 신뢰 신경망 (Deep Belief Network, DBN)

심층 신뢰 신경망은 제한된 볼츠만 머신 (Restricted Boltzmann Machine)들의 결합으로 이루어진다. 심층 신뢰 신경망에는 각 층(Layer)이 있고, 각 층에는 데이터가 존재하는 가장 작은 단위인 노드(Node)가 있다. 각 노드간의 연결은 가중치(Weight)라고 한다. 이 때 인접한 두 레이어를 하나의 제한된 볼츠만 머신이라고 한다. (Fig. 3)은 본 연구에서 제안하는 심층 신뢰 신경망의 구조이다. 입력층은 6 \* 20의 120개의 노드로 구성된다. 데이터는 한 시점에 가속도 센서 데이터 3개와 자이로스코프 센서 데이터 3개가 수집되며, 시간의 변화에 따른 흐름을 인식시키기 위하여 20회를 묶어서 하나의 구간으로 구성하였다. 은닉층은 제한된 볼츠만 머신으로 구성하는데, 은닉층 전체를 하나의 제한된 볼츠만 머신으로 구성하지 않고, 구간을 나누어 구성하였다. 첫 번째

은닉층은 입력층의 크기를 달리하여, 입력층의 구간 크기를 5인 네 개의 제한된 볼츠만 머신과 입력층의 구간 크기를 10으로 가지는 2개의 제한된 볼츠만 머신, 입력층의 구간 크기를 20으로 가지는 1개의 제한된 볼츠만 머신으로 모두 7개의 제한된 볼츠만 머신을 구성하였다. 두 번째 은닉층은 첫 번째 은닉층의 출력을 입력으로 사용하는데, 입력층의 구간 크기를 5인 4개의 제한된 볼츠만 머신들을 결합한 제한된 볼츠만 머신과 입력층의 구간 크기를 10으로 가지는 2개의 제한된 볼츠만 머신들을 결합한 제한된 볼츠만 머신, 그리고 입력층의 구간 크기를 20으로 가지는 제한된 볼츠만 머신의 3개의 제한된 볼츠만 머신으로 구성하였다. 마지막 세 번째 은닉층은 두 번째 은닉층의 3개의 제한된 볼츠만 머신들을 결합하여 1개의 제한된 볼츠만 머신으로 구성하였다. 출력층은 인간의 행동을 구분하는 6개의 노드로 구성하였다.

심층 신뢰 신경망의 학습은 사전 학습과 파인튜닝 단계로 구분된다. 사전 학습 단계에서는 심층 신뢰 신경망의 모든 가중치들의 초기값을 구성하는 단계로, 주어진 입력 데이터에 대해 비지도 학습 (Unsupervised Learning)의 훈련을 수행하는 것으로 단계별로 제한된 볼츠만 머신을 구성해서 학습하였다. 초기값을 구성한 후 목표값에 잘 맞도록 심층 신뢰 신경망을 학습하는 과정을 파인 튜닝이라고 하며, 지도 학습 (Supervised Learning)의 훈련을 수행하는 것으로 오류 역전파 알고리즘으로 학습하였다.

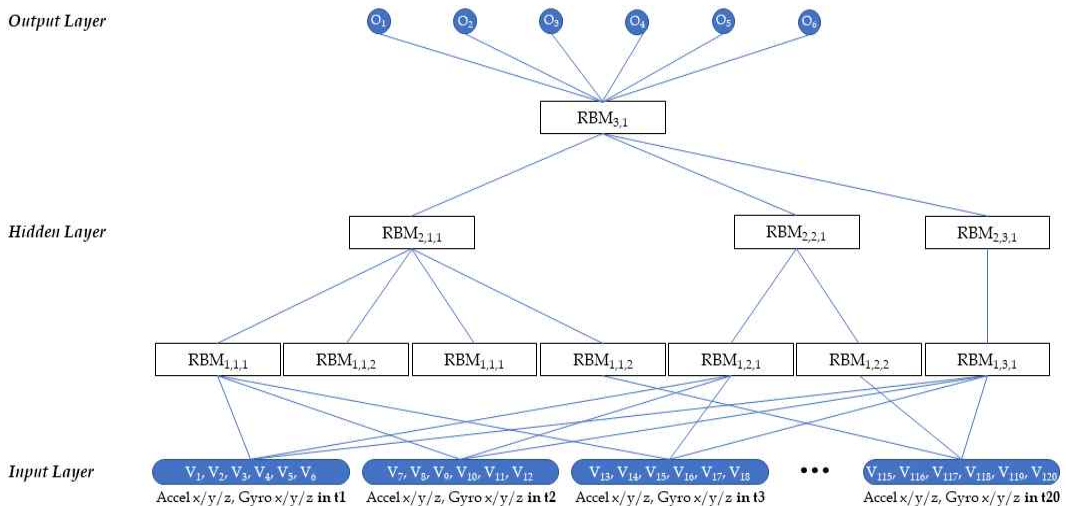


Fig. 3. The proposed DBN architecture.

### 3.4 제한된 볼츠만 머신 (Restricted Boltzmann Machine, RBM)

제한된 볼츠만 머신은 차원 감소, 분류, 선형 회귀 분석, 협업 필터링 등에 사용할 수 있는 알고리즘으로 제프리 힌튼(Geoff Hinton)이 제안하였다. 제한된 볼츠만 머신은 자체적으로도 사용할 수 있지만, 다른 신경망의 학습을 돕기 위해서도 사용된다. 제한된 볼츠만 머신은 가시층(Visible Layer)이라고도 불리는 입력층 1개와 은닉층 1개로 이루어진 단층 신경회로망이다. 제한된 볼츠만 머신은 입력층의 입력 노드 ( $v_i$ )와 은닉층의 은닉 노드( $h_j$ )를 연결하는 가중치 매트릭스 ( $W = w_{i,j}$ )로 구성되며, 입력층과 은닉층에는 바이어스(bias) 가중치  $a_i$ 와  $b_j$ 가 존재한다. 이 구성의 에너지 함수는 수식 (1)과 같다.

$$E(v, h) = -a^T v - b^T h - v^T W h \quad (1)$$

은닉 노드와 입력 노드 간의 확률 분포는 수식 (2)의 에너지 함수를 사용하여 정의된다.

$$P(v, h) = \frac{1}{Z} e^{-E(v, h)} \quad (2)$$

여기서,  $Z$ 는  $Z = \sum_{x, y} e^{-E(x, y)}$ 로 정의되는 파티션 함수(partition function)로 확률 분포의 합을 1이 되게끔 한다. 입력 층의 확률은 은닉 층의 구성의 합이 된다.

$$P(v) = \frac{1}{Z} \sum_h e^{-E(v, h)} \quad (3)$$

제한된 볼츠만 머신의 입력 노드들은 서로 독립적이므로 은닉 노드들의 활성화에 의한 입력 노드의 활성 여부는 서로 독립적이고, 마찬가지로, 입력 노드들의 활성화에 의한 은닉 노드의 활성 여부는 서로 독립적이다. 즉,  $m$ 개의 입력 노드와  $n$ 개의 은닉 노드가 있을 때 입력층의 조건부 확률은 수식 (4)와 같이 모든 입력 노드들의 조건부 확률의 곱으로 계산되고, 은닉층의 조건부 확률도 수식 (5)와 같이 모든 은닉 노드들의 조건부 확률의 곱으로 계산된다.

$$P(v|h) = \prod_{i=1}^m P(v_i|h) \quad (4)$$

$$P(h|v) = \prod_{j=1}^n P(h_j|v) \quad (5)$$

제한된 볼츠만 머신의 학습은 학습 집합  $V$ 의 조건부 확률을 최대화하기 위한 방향으로 이루어진다. 이

를 위해 많이 사용되는 알고리즘은 힌튼(Hinton)의 CDA (contrastive divergence) 알고리즘이다[12]. 학습은 기브스 표집(Gibbs Sampling)과 기울기 하강(Gradient Descent) 기법을 사용하며, CDA 알고리즘에 의해 가중치와 바이어스들은 다음과 같이 계산된다. 수식 (6), (7), (8)을 이용하여 가중치와 두 개의 바이어스를 수정하는 작업을 반복한다.

$$\Delta W = \alpha(vh^T - v'h'^T) \quad (6)$$

$$\Delta a = \alpha(v - v') \quad (7)$$

$$\Delta b = \alpha(h - h') \quad (8)$$

은닉층의 제한된 볼츠만 머신을 하나로 구성한 것이 아니라, 데이터 구간에 따라 다양하게 구성하였다. (Fig. 3)과 같이 입력층은 한 시점의 센서로부터 6개의 입력을 받고, 20개의 시점을 한 번에 입력으로 사용하기 때문에 120개의 입력 노드를 가지고 있다. 첫 번째 은닉층의 RBM<sub>1,1,1</sub> ~ RBM<sub>1,1,4</sub>는 5개의 시점의 센서 데이터가 입력으로 들어오도록 구성하여, 입력층은 30개의 입력 노드로, 은닉층은 12개의 은닉 노드로 구성하였다. 첫 번째 은닉층의 RBM<sub>1,2,1</sub> ~ RBM<sub>1,2,2</sub>는 10개의 시점의 센서 데이터를 입력으로 들어오도록 구성하여, 입력층은 60개의 입력 노드로, 은닉층은 24개의 은닉 노드로 구성하였다. 첫 번째 은닉층의 RBM<sub>1,3,1</sub>은 20개의 시점의 센서 데이터를 입력으로 구성하여, 입력층은 120개의 입력 노드로, 은닉층은 48개의 은닉 노드로 구성하였다. 두 번째 은닉층의 RBM<sub>2,1,1</sub>은 RBM<sub>1,1,1</sub> ~ RBM<sub>1,1,4</sub>의 은닉층을 입력층으로 받아들이고, RBM<sub>2,2,1</sub>은 RBM<sub>1,2,1</sub> ~ RBM<sub>1,2,2</sub>의 은닉층을 입력층으로 받아들이며, RBM<sub>2,3,1</sub>은 RBM<sub>1,3,1</sub>의 은닉층을 입력층으로 받아들이므로, 세 RBM은 모두 입력층은 48개의 입력 노드로 설정하고, 은닉층은 12개의 은닉 노드로 구성하였다. 세 번째 은닉층의 RBM<sub>3,1,1</sub>은 RBM<sub>2,1,1</sub> ~ RBM<sub>2,3,1</sub>의 은닉층을 입력층으로 받아들이므로, 입력층은 36개의 입력 노드로 설정하고, 은닉층은 전체 모델의 출력이 되므로, 인간의 행동을 분류하는 6개로 구성하였다. 사전 학습 단계는 제한된 볼츠만 머신들을 학습하는 것으로, 제한된 볼츠만 머신의 학습은 첫 번째 은닉층의 일곱 개의 RBM들을 학습하고, 안정이 된 후 두 번째 은닉층의 세 개의 RBM들을 학습하고, 마지막으로 세 번째 은닉층의 RBM을 학습하였다.

3.5 파인튜닝(Fine Tuning)

네트워크의 모든 제한된 볼츠만 머신들의 초기 가중치를 설정하는 사전 학습 단계를 거치고 난 후 가중치들과 바이어스들을 수정하는 작업인 파인튜닝을 수행하였다. 이 과정에서는 오류 역전파 (Error Back Propagation) 알고리즘으로 각 시점별로 인간의 행동을 목표값으로 설정하여 목표값과 신경회로망의 결과를 비교하여 학습하는 지도 학습을 하였다.

오류 역전파 알고리즘에서 가중치는 수식 (9)와 같은 확률적 경사 하강법(stochastic gradient descent)을 이용하여 갱신한다.

$$w_{ij}(t+1) = w_{ij}(t) - \alpha \frac{\delta C}{\delta w_{ij}} \tag{9}$$

여기서,  $w_{ij}$ 는 뉴런  $i$ 와 뉴런  $j$  사이의 가중치이고,  $\alpha$ 는 학습률이며,  $C$ 는 비용 함수 (cost function)이다.

$\frac{\delta C}{\delta w_{ij}}$ 는 가중치에 대한 비용 함수의 기울기로서

$\frac{\delta C}{\delta w_{ij}} = o_i \delta_j$ 으로 계산된다.  $o_i$ 는 뉴런  $i$ 의 출력이고,  $\delta_j$ 는 수식 (10)과 같이 계산된다.

$$\delta_j = \begin{cases} (o_j - t_j) o_j (1 - o_j) & \text{if } j \text{는 출력뉴런} \\ (\sum \delta_i w_{ij}) o_j (1 - o_j) & \text{if } j \text{는 은닉뉴런} \end{cases} \tag{10}$$

4. 실험 및 결과

4.1 실험 데이터

실험데이터는 Anguita 등[6]에 의해서 공개된 스마트폰에서 3D 가속도(Accelerometer) 센서, 3D 자이로스코프(Gyroscope) 센서의 데이터를 50Hz (20 ms 당 1회)로 측정된 센서 데이터를 대상으로 하였다. 가속도 센서와 자이로스코프 센서의 원본 데이터 (Raw Data) 6개를 입력으로 하였다. 센서에서 측정된 현재의 값만을 입력으로 하지 않고, 현재 시점을 포함하여 20회 측정된 데이터, 즉, 0.4초 동안 누적된 값을 입력으로 하여 구간의 변화를 반영할 수 있도록 하였다. 학습 데이터는 7352개이고, 테스트 데이터는 2947개이다. 연속적인 변화를 반영하기 위해 20회 측정된 데이터를 사용하기 때문에, 학습 데이터는 7333개, 테스트 데이터는 2928개이다.

데이터 집합은 걷기(walking), 계단 오르기(upstairs), 계단 내려가기(downstairs), 앉기(sitting), 일어서기(standing), 눕기(laying)의 6가지 행동이 기록되어 있다. 전체 데이터 중 일부 데이터를 인간의 행동을 기준으로 살펴보면 (Fig. 4)와 같다. 걷기, 계단 오르기, 계단 내려가기와 같이 움직임이 큰 경우는 센서 데이터의 변동폭이 크고, 앉기, 일어서기, 눕기 등 움직임이 작은 경우는 센서 데이터의 변동폭이 작은 특성을 가지고 있다.

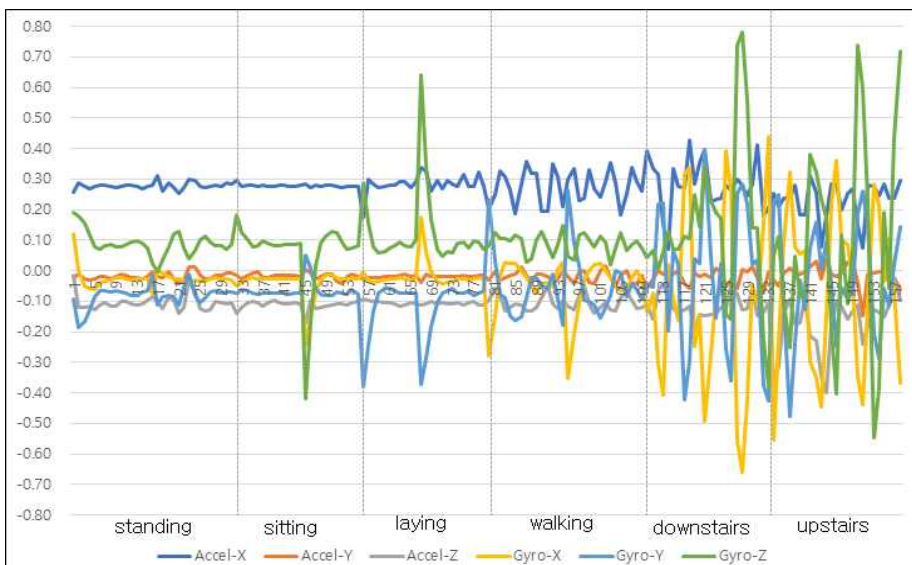


Fig. 4. Sample of 6 Sensor Data.

Table 1. Accuracy by learning rate

Learning rate	Accuracy	Learning rate	Accuracy
0.1	95.317 %	0.009	99.118 %
0.09	95.148 %	0.008	99.593 %
0.08	95.521 %	<b>0.007</b>	<b>99.661 %</b>
0.07	96.064 %	0.006	99.389 %
0.06	96.878 %	0.005	99.084 %
0.05	97.218 %	0.004	98.846 %
0.04	97.523 %	0.003	98.948 %
0.03	97.727 %	0.002	99.050 %
0.02	98.914 %	0.001	98.914 %
0.01	98.982 %		

4.2 학습 결과

학습율(Learning Rate)에 따른 제안하는 테스트 데이터에 대한 심층 신뢰 신경망의 성능의 변화는 <Table 1>과 같다. <Table 1>과 같이 학습율이 0.04 이하일 때 테스트 데이터에 대한 정확도(Accuracy)가 좋았다. 학습율이 0.007일 때 테스트 데이터에 대해 가장 우수한 성능을 보였으며, 그때의 정확도는 99.661%였다.

4.3 제안하는 모델의 성능

인간의 행동 인식에 대한 다른 알고리즘들과 제안

하는 DBN 기반 모델의 성능을 비교하였다. (Fig. 5)는 동일한 데이터 집합[6]에 대해 제안하는 DBN 모델의 성능과 SVM[6], AdaBoost[8], IBk[9], Convnet[11] 알고리즘과 인간의 행동별 정밀도(Precision)를 비교하였다.

정밀도는  $TP/(TP+FP)$ 로 나타내며, 맞춘 문제 중 정답의 개수에 대한 비이다. TP는 True Positive로 True라고 예측했는데, 실제로 True인 경우이다. FP은 False Positive로 True로 예측했는데, 실제로 False인 경우이다. 정밀도가 1에 가까울수록 성능이 좋은 것이다. 움직임의 변화가 작은 일어서기, 앉기, 눕기의 경우에는 구간의 범위가 작고 크에 따라 다른 변화를 보이기 때문에 다양한 구간으로 제한된 볼츠만 머신을 구성했을때 정밀도가 가장 우수하였다. 움직임의 변화가 큰 걷기, 계단 오르기, 계단 내려가기의 경우에는 여러 구간의 범위에서 항상 변동이 크기 때문에 다양한 구간으로 제한된 볼츠만 머신을 구성해도 평균정밀도는 0.7 % 정도 낮았다. 하지만 인간의 행동별 최대 정밀도와 최소 정밀도의 차는 SVM은 9.913 %, AdaBoost는 1.518 %, IBk는 15.8 %, Convnet은 13.3 % 인데 비하여, 제안하는 DBN 모델은 0.955 %로 다양한 구간을 반영하기 때문에 안정성이 가장 우수하였다. 제안하는 DBN 모델은 특정 행동만을 인식하는데 치우치지 않고, 전체적으로 안정적인 결과를 보였다.

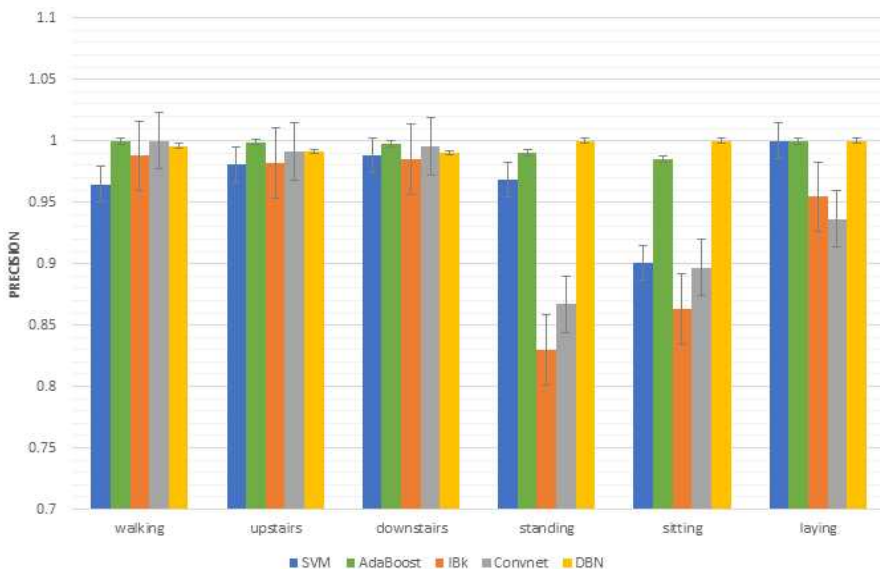


Fig. 5. Human activity recognition precision of SVM, AdaBoost, IBk, Convnet and proposed DBN model.



## 5. 결 론

본 논문은 스마트폰의 가속도 센서와 자이로스코프 센서의 데이터를 이용하여 딥러닝(Deep Learning) 알고리즘의 하나인 심층 신뢰 신경망을 구성하여 인간의 행동을 인식하는 시스템을 위한 기초연구이다. 심층 신뢰 신경망을 구성할 때 은닉층의 제한된 볼츠만 머신을 하나로 구성하는 것이 아니라 계층적으로 구성하는데, 연속된 데이터를 받아서 여러 간격의 구간 특성을 반영하여 각각의 제한된 볼츠만 머신을 구성함으로써 짧은 간격과 긴 간격을 확인할 수 있도록 구성하여 보다 정확하게 행동을 인식할 수 있었다. 움직임의 변화가 큰 행동은 긴 간격구간에서의 관찰로 인식 성능을 높일 수 있고, 움직임의 변화가 작은 행동은 짧은 간격구간에서의 관찰에서도 높은 인식 성능을 보일 수 있다. 따라서 시간의 간격구간을 달리하여 제한된 볼츠만 머신을 구성함으로써 인식의 성능을 높이고 안정적인 결과를 보일 수 있었다.

본 논문에서 제안하는 모델은 딥러닝의 가장 기본이 되는 심층 신뢰 신경망(Deep Belief Network)을 기반으로 하기 때문에 사전에 특징 추출이 없이 원 데이터를 그대로 사용할 수 있는 장점을 가진다. 심층 신뢰 신경망은 전체 입력을 하나의 데이터로 가정하기 때문에, 시계열 특성을 가지는 데이터에서 구간의 특성을 반영하지 못하는 단점이 있다. 본 논문에서는 제한된 볼츠만 머신을 구성할 때에 누적된 시간의 데이터를 사용하고 이들에 대한 시간의 간격을 여러 가지로 하여 심층 신뢰 신경망을 구성하여 이러한 단점을 극복하였다. 하지만 이 방식은 네트워크의 구조가 복잡하기 때문에 복잡도 개선을 위한 연구가 추가적으로 필요하다.

## REFERENCE

[1] J. Wang, Z. Liu, Y. Wu, and J. Yuan, "Learning Actionlet Ensemble for 3D Human Action Recognition," *IEEE Transactions on Pattern Analysis Machine Intelligence*, Vol. 36, pp. 914-927, 2013.

[2] P. Li, Y. Wang, Y. Tian, T.S. Zhou, and J.S. Li, "An Automatic User-Adapted Physical Activity Classification Method Using Smart-

phones," *IEEE Transactions on Biomedical Engineering*, Vol. 64, No. 3, pp. 706-714, 2017.

[3] J. Wen and Z. Wang, "Sensor-based Adaptive Activity Recognition with Dynamically Available Sensors," *Neurocomputing*, Vol. 218, pp. 307-317, 2016.

[4] J.I. Choi and H.S. Yong, "Activity Data Modeling and Visualization Method for Human Life Activity Recognition," *Journal of Korea Multimedia Society*, Vol. 15, No. 8, pp. 1059-1066, 2012.

[5] M.C. Lee and S.B. Cho, "Accelerometer-based Gesture Recognition Using Hierarchical Recurrent Neural Network with Bidirectional Long Short-Term Memory," *Journal of Korean Institute of Information Scientists and Engineers : Software and Applications*, Vol. 39, No. 12, pp. 1005-1011, 2012.

[6] D. Anguita, A. Ghio, L. Oneto, X. Parra, and J.L. Reyes-Ortiz, "A Public Domain Dataset for Human Activity Recognition Using Smartphones," *Proceeding of European Symposium on Artificial Neural Networks, computational Intelligence and Machine Learning*, pp. 437-442, 2013.

[7] A. Stisen, H. Blunck, S. Bhattacharya, T.S. Prentow, M.B. Kjærgaard, A. Dey, et al., "Smart Devices are Different: Assessing and Mitigating Mobile Sensing Heterogeneities for Activity Recognition," *Proceeding of 13th Association for Computing Machinery Conference on Embedded Networked Sensor Systems*, pp. 127-140, 2015.

[8] A. Reiss, G. Hendeby, and D. Stricker, "A Competitive Approach for Human Activity Recognition on Smartphones," *Proceeding of European Symposium an Artificial Neural Networks*, pp. 455-460, 2013.

[9] G. Chetty, M. White, and F. Akther, "Smart Phone Based Data Mining For Human Activity Recognition," *Procedia Computer Science*, Vol. 40, pp. 1181-1187, 2015.



- [10] C.A. Ronao and S.B. Cho, "Human Activity Recognition Using Smartphone Sensors with Two-stage Continuous Hidden Markov Models," *Proceeding of 10th International Conference on Natural Computation*, pp. 686-691, 2014.
- [11] C.A. Ronao and S.B. Cho, "Human Activity Recognition with Smartphone Sensors Using Deep Learning Neural Networks," *Expert Systems with Applications*, Vol. 59, pp. 235-244, 2016.
- [12] W. Jiang and Z. Yin, "Human Activity Recognition Using Wearable Sensors by Deep Convolutional Neural Networks," *Proceeding of the 23rd Association for Computing Machinery International Conference on Multimedia*, pp. 1307-1310, 2015.
- [13] H.J. Lee, "Human Activity Recognition Using Multi-temporal Neural Networks," *Journal of Digital Contents Society*, Vol. 18, No. 3, pp. 151-159, 2017.
- [14] Y.C. Lee and C.W. Lee, "Motion Recognition of Smartphone Using Sensor Data," *Journal of Korea Multimedia Society*, Vol. 17, No. 12, pp. 1437-1445, 2014.



### 이 현 진

1996년 순천향대학교 전산학과  
공학사

1998년 연세대학교 대학원 컴퓨  
터과학과 공학석사

2002년 연세대학교 대학원 컴퓨  
터과학과 공학박사

현재 숭실사이버대학교 ICT공학부 부교수  
관심분야: 이러닝, 머신러닝, 빅데이터 분석