

# LSTM을 이용한 사용자 활동유형 및 인식기술 개발

김영균<sup>1</sup> · 김원종<sup>2</sup> · 이석원<sup>1</sup>

<sup>1</sup>아주대학교 · <sup>2</sup>한국전자통신연구원

## Development of user activity type and recognition technology using LSTM

Young-kyun Kim<sup>1</sup> · Won-jong Kim<sup>2</sup> · Seok-won Lee<sup>1</sup>

<sup>1</sup>Ajou University · <sup>2</sup>ETRI

E-mail : rladudrbs06@naver.com / wjkim@etri.re.kr / leesw@ajou.ac.kr

### 요 약

인간의 활동은 척추 옆굽음증, 골반 뒤틀림과 같은 개개인의 신체적 특징부터 기쁨, 분노, 슬픔 등의 감정들까지 다양한 요인들에 영향을 받는다. 하지만 이러한 동작의 특성은 오랜 시간에 걸쳐서 변화하며, 단기적으로 행동의 특성은 크게 변하지 않는다. 사람의 활동 데이터는 시간 흐름에 따라서 변화하는 시계열 적 특징과 각 행동별로 일정한 규칙성을 갖는다. 본 연구에서는 시계열 적 특징을 다루기 위한 순환신경망의 한 종류인 LSTM을 활동유형을 인식하는 기술에 적용하였으며, 측정시간과 LSTM 모델의 구성요소들에 대한 파라미터 최적화로 활동유형의 인식률을 개선하였다.

### ABSTRACT

Human activity is influenced by various factors, from individual physical features such as vertebral flexion and pelvic distortion to feelings such as joy, anger, and sadness. However, the nature of these behaviors changes over time, and behavioral characteristics do not change much in the short term. The activity data of a person has a time series characteristic that changes with time and a certain regularity for each action. In this study, we applied LSTM, a kind of cyclic neural network to deal with time - series characteristics, to the technique of recognizing activity type and improved recognition rate of activity type by measuring time and parameter optimization of components of LSTM model.

### 키워드

Neural network, LSTM, Activity, Recognition

## I. 서 론

인간의 활동은 개개인의 신체구조, 건강, 심리상태 등의 요인들로부터 그 특성이 나타난다. 여자와 남자, 노인과 아이, 행복한 사람과 우울한 사람 등 그룹 간 특징에 따라 자리에 앉기, 걷기, 뛰기 등 단순한 행동이지만 미세한 행동의 차이는 존재한다. 이렇게 다양하게 정의되는 행동특징을 분석하고 연구하는 많은 시도가 이루어지고 있다[1][2]. 선행연구들에서는 예측모델을 제작한 후 사람의 행동을 예측, 평가하고 있다. 본 연구는 선행연구를 발전시켜 측정시간, 예측모델의 구조에 변화를 주어 인식률을 개선하였다.

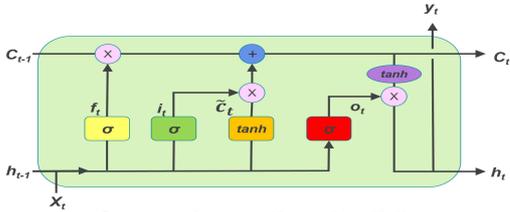
본 연구에서 다루고자 하는 사람의 행동 특성은

시간 흐름에 따라서 변화하는 특징이 있다. 이러한 시계열 적 특징을 효과적으로 다루는 다양한 방법들중에서 본 연구는 RNN(Recurrent Neural Network)의 한 종류인 LSTM(Long Short Term Memory)을 사용해 그 문제를 해결하고 있다. 본 연구에서는 LSTM을 적용한 예측모델로 사람 활동을 비교, 분석했다.

## II. 활동인식 모델개발

### 2. 활동인식 모델

그림 1. LSTM의 내부구조[3]



LSTM은 RNN에서 파생된 한 종류로서, RNN의 단점인 장기 의존성 문제를 해결하기 위해 고안된 방법이다[4]. 이를 위해 하이퍼볼릭 탄젠트(Tanh) 함수만 사용하는 기존 RNN에서 [그림 1]과 같이 LSTM 블록 내부에 3가지의 게이트(forget, input, output)를 추가해 그 문제를 해결하고 있다.

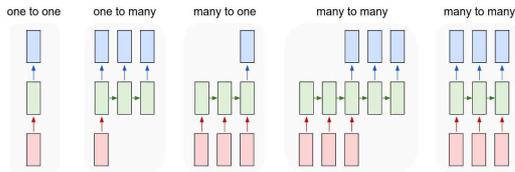


그림 2. LSTM의 구조[5]

LSTM에는 [그림 2]와 같이 다양한 구조가 있다. 각각의 LSTM 구조는 풀고자 하는 문제에 따라 다르게 구조화된다. 하나의 입력값을 받아 다수의 출력값을 얻는 문제에서는 일대 다수(One to Many) 구조, 다수의 입력과 다수의 출력이 요구되는 경우 다수대 다수(Many to Many) 구조로 해결한다.

본 연구는 사람의 연속된 활동을 입력으로 받아 한가지 활동을 결정하는 문제에 해당한다. 연속된 활동은 시간 흐름에 따른 다수의 입력값으로서 그 크기는 특정 활동의 측정시간에 따라 결정된다. 따라서 연구는 다수대 일(Many to one)의 LSTM 구조를 적용하였다.

### 2.1 활동 데이터

본 연구에서 사용하는 데이터는 Wireless Sensor Data Mining Lab에서 제공하는 WISDM 데이터셋을 사용하였다[6]. WISDM 데이터는 총 36명의 피실험자를 대상으로 수집되었고 6가지 활동(걷기, 조깅, 계단 올라가기, 계단 내려가기, 앉기, 서 있기)을 실험했다. 피실험자들은 각자 청바지 앞주머니에 안드로이드 스마트폰을 넣고 내장된 가속도센서(Accelerometer)를 사용해 20Hz 단위로 초당 20개의 센서값 데이터를 수집했다. 가속도센서로 부터 활동 방향에 대한 X, Y, Z의 가속도 값을 얻었고, 센서값들로부터 6가지 활동인식을 진행했다. 전반적인 WISDM 데이터 설명은 [표 1]과 같다.

※ null value의 개수는 x, y, z 센서값 모두 0일 때를 의미

실험자 수	활동 수	Feature 수	Row 수	NULL value 수
36	6	3	1,098,205	12,843

표 1. WISDM 데이터

### 2.2 데이터 전처리

좋은 결과를 얻기 위해서는 잘 정제된 데이터를 입력해

야 하므로 데이터의 구성과 특징을 사전에 파악하는 과정이 필요하다. [표 1]에서 WISDM 데이터의 결측값은 12,843개로 전체 Row 수(1,098,205개)의 약 1.1%의 적은 부분을 차지하고 있다. 일부 수집되지 못한 결측치를 센서의 물리적 결함으로 판단하고, 보간(Interpolate) 함수를 사용하여 처리하였다.

또한, 모델 제작을 위해 WISDM 데이터 중 70%는 훈련용, 30%는 검증용으로 랜덤하게 나누었다.

### 2.3 파라미터 최적화

Segment 크기	LSTM 층 수	활성화 함수	손실 함수	옵티마이저
50, 100, 200	2, 4, 6	tanh	L2 loss function	Adam

표 2. LSTM 모델의 구성요소

LSTM 모델의 요소들은 [표 2]와 같고 Tensorflow에서 진행했다. 연구는 모델학습을 위해 데이터를 일정한 크기의 Segment 단위로 구분한 후 학습했다. 이는 LSTM에서 데이터를 활용하여 인식모델을 학습시킬 것 인지에 해당한다.

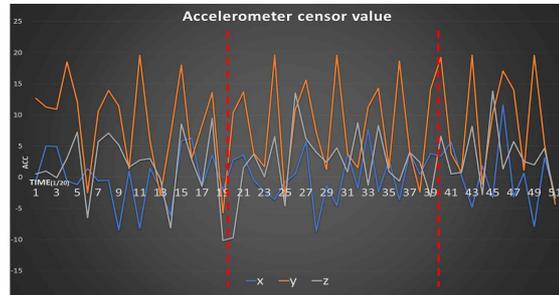


그림 3. 가속도센서 값

[그림 3] 가속도센서에서 X, Y, Z 값을 추출한 그래프의 한 예를 보인 것이다. 센서에서 20Hz 단위로 값을 측정하므로 가로축에서 한 칸은 1/20초가 된다. 그래프에서 빨간색 점선은 Segment 크기를 20개씩 자른 것으로 1초를 뜻한다. 즉, 모델학습을 위해 Time step 값을 1초 단위로 측정했다는 의미이다. 본 연구에서는 Segment 크기를 50, 100, 200개씩 나누었는데, 각각 2.5초, 5초, 10초의 측정시간과 일치한다.

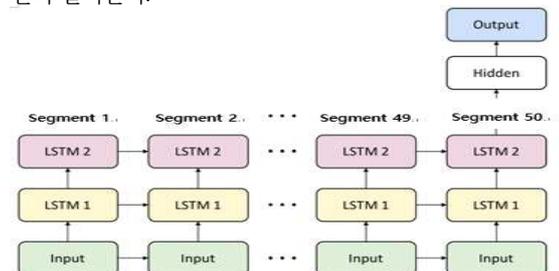


그림 4. LSTM 네트워크구조

[그림 4]는 [그림 2]에서 설명한 다수대 일(Many to

one)의 LSTM 네트워크구조를 본 연구에 적용한 모델이다. 그림에서는 두 개의 LSTM 층과 Segment의 크기가 50 (2.5초 측정)일 때를 보여주고 있다.

모델의 예측력은 LSTM 층의 개수에 따라 바뀌고 영향을 받는다. 이는 데이터의 상태에 따라 LSTM 네트워크구조와 층의 개수가 결정된다. 데이터의 개수가 많고 고차원일 경우 일반적으로 신경망 구조를 깊게 하여 LSTM 층을 깊게 쌓는다. 본 연구에서 사용된 데이터의 Row 수는 1,098,205개, Feature 수는 3개로 구성되어있다. 비교적 저차원 데이터에 해당하는데, 본 연구에서는 LSTM 층의 개수를 차례대로 2, 4, 6개씩 구조화하여 진행했다.

### III. 실험 결과

Segment	Activity(%)	Down stairs	Jog	Sit	Stand	Up stairs	Walk
1		0.77	97.1	0.15	0.2	0.66	1.11
2		0.01	99.8	0	0.08	0.11	0
3		30.1	44.9	0	0	24.8	0
....							
2043		0.16	87.9	0	0.06	11.8	0.02
2044		0	99.9	0	0	0	0
평균		0.72	97.6	0	0	0.72	0.9
활동 별 예측한 Segment 수		13	2007	0	0	8	16
활동 별 예측한 Segment 수(%)		0.63	98.1	0	0	0.39	0.78

표 3. Jog Segment의 각 활동 별 예측 (Segment size=50)

[표 3]는 Jog에 대한 Segment 크기가 50일 때의(활동을 2.5초 동안 측정) 각 Segment별 인식률 결과를 나타낸 표이다. Segment 크기가 50일 때 Jog의 Segment 수는 총 2,044개로 구성되었다. [표 3]에서 확률값 들은 각각의 Segment들이 해당하는 활동으로 인식하는 확률을 의미하는 것으로, 한 Segment의 6가지 확률값을 더하면 100%가 된다. 예를 들어, 첫 번째 Jog Segment의 확률값 (0.77, 97.1, 0.15, 0.2, 0.66, 1.11)을 더하면 약 100%가 된다. 또한, 확률값 들 중 두 번째 항목인 Jog가 97.1%로 6개 값 중 가장 높아 해당 Segment는 Jog 활동으로 인식한 것이다.

[표 3]에서 Jog를 예측한 Segment들의 확률들의 평균값은 97.6%로 대다수의 Segment 들이 Jog로 인식되었다. 이는 Jog Segment에 대해 진행했으므로 당연히 Jog를 예측해야 하지만, 일부 Segment들에 대해서 다른 활동으로 인식한 경우도 있다. [표 3]에서 전체 Segment 개수는 2044개 중 Jog를 예측한 Segment 개수는 2007개로 인식률은 98.17%이다.

The number of LSTM layer = 2							
활동 별 예측한 Segment 수		Down stairs	Jog	Sit	Stand	Up stairs	Walk
Downstairs	Seg 50	70.6	4.01	0	0	13.1	12.2
	Seg 100	61.8	3.52	0	0	19.2	15.3
	Seg 200	56.3	1.37	0	0	25.5	16.7
Jog	Seg 50	0.63	98.1	0	0	0.39	0.78
	Seg 100	0.57	97.7	0	0	1.06	0.57
	Seg 200	0.77	94.9	0	0	1.54	2.71
Sit	Seg 50	0	0	99.8	0	0.01	0
	Seg 100	0	0	97	1.74	1.16	0
	Seg 200	0	0	96.9	1.02	1.02	1.02
Stand	Seg 50	0.34	0	0.11	97.9	0.68	0
	Seg 100	0	0	0.62	98.7	0.62	0
	Seg 200	1.42	0	2.85	95.7	0	0
Upstairs	Seg 50	15.9	7	0.41	0.41	64.6	11.4
	Seg 100	15.3	6.5	0.25	0.75	63.4	13.6
	Seg 200	18.8	8.93	0.52	0	59.6	12.1
Walk	Seg 50	1.94	0.77	0.03	0	1.86	95.3
	Seg 100	1.87	0.57	0	0	2.61	94.8
	Seg 200	1.6	0.96	0.16	0	2.56	94.7

표 4. Segment 크기 변화에 따른 6가지 활동 예측 (LSTM layer=2)

The number of LSTM layer = 4							
활동 별 예측한 Segment 수 (%)		Down stairs	Jog	Sit	Stand	Up stairs	Walk
Downstairs	Seg 50	68.2	2.83	0	0	18.3	10.4
	Seg 100	38.5	3.17	0.25	0.75	36.7	20.5
	Seg 200	54.1	3.66	0.12	0.07	18.4	23.5
Jog	Seg 50	0.83	96.5	0.01	0	1.68	0.9
	Seg 100	0.37	96.9	0.01	0.01	1.47	1.17
	Seg 200	0.99	95.5	0.01	0	1.26	2.19
Sit	Seg 50	0.19	0.05	98	0.77	0.83	0.03
	Seg 100	0.06	0.08	97.1	1.41	1.19	0.11
	Seg 200	1.3	0.09	94.5	2.95	1.09	0
Stand	Seg 50	0.43	0.01	1.57	95.4	2.25	0.21
	Seg 100	0.04	0.01	0.7	98	1.11	0.1
	Seg 200	1.62	0	0.48	96.5	1.48	0.02
Upstairs	Seg 50	12.4	6.47	0.55	0.32	66.3	13.7
	Seg 100	7.38	3.82	0.28	0.62	72.6	15.2
	Seg 200	28.8	7.02	0.93	0.13	36.5	26.6
Walk	Seg 50	1.92	0.68	0.01	0	2.69	94.6
	Seg 100	0.24	0.62	0	0	3.06	96
	Seg 200	2.89	0.52	0.2	0	1.49	94.8

표 5. Segment 크기 변화에 따른 6가지 활동 예측 (LSTM layer=4)

The number of LSTM layer = 6							
활동 별 예측한 Segment 수 (%)		Down stairs	Jog	Sit	Stand	Up stairs	Walk
Downstairs	Seg 50	61.2	1.88	0.15	0.06	25	11.5
	Seg 100	37	5.4	0.34	0.29	34.4	22.4
	Seg 200	71.6	0.78	0.03	0.05	22.2	5.18
Jog	Seg 50	1.27	93.8	0.03	0	4	0.88
	Seg 100	0.07	98.4	0.01	0	1.16	0.31
	Seg 200	2.53	93.3	0	0	2.82	1.32
Sit	Seg 50	0.22	0.06	98.1	0.67	0.71	0.15
	Seg 100	0.02	0.03	97.9	1.02	0.92	0
	Seg 200	1.09	0.03	95.1	1.55	2.1	0.02
Stand	Seg 50	0.24	0.04	2.21	95.9	1.26	0.26
	Seg 100	0.13	0.01	2.34	96.4	1.01	0
	Seg 200	0.02	0.06	1.2	98.5	0.1	0.01
Upstairs	Seg 50	14.7	3.12	0.62	0.45	66.1	14.9
	Seg 100	7.48	5.83	0.27	0.82	68.9	16.1
	Seg 200	15.4	6.15	0.05	0.85	66	11.4
Walk	Seg 50	2.31	0.47	0.01	0.01	3.92	93.2
	Seg 100	0.54	0.29	0	0	1.92	97.2
	Seg 200	3.06	1.28	0	0	4.94	90.7

표 6. Segment 크기 변화에 따른 6가지 활동 예측 (LSTM layer=6)

[표 4], [표 5], [표 6]는 LSTM 층 개수를 차례대로 2, 4, 6개씩 구성해 가면서 6가지 활동을 예측한 결과표다. 표 안의 색깔은 초록색에 가까워질수록 예측력이 높은 것이고, 빨간색은 그 반대이다. [표 4], [표 5], [표 6]에서 세로축은 실제 활동, 가로축은 [표 3]의 '활동 별 예측한 Segment 수 (%)'와 그 의미가 일치한다. 3개의 표에서 공통 적으로 대각선은 초록색계열이 짙은 것을 볼 수 있는데, Confusion Matrix의 Accuracy와 같이 전체 중 올바르게 예측한 것들로 초록색계열이 짙으면 예측모델의 인식률이 높은 것을 뜻한다.

[표 4]에서 Stand를 제외한 나머지 5가지 활동에 대해 Segment의 크기가 50에서 200으로 증가할수록 확률값이 감소하는 것을 볼 수 있다. 즉, 측정시간이 2.5초에서 10초로 길어질수록 알맞은 활동을 예측할 확률의 감소를 뜻한다. 그러나, Stand는 Segment 크기가 50에서 100으로 증가 할 때만 예측확률이 97.9%에서 98.7%로 오히려 증가하는 것을 확인할 수 있다. LSTM 층이 2개인 예측모델에서 Stand를 인식할 때 측정시간이 2.5초보다는 5초로 측정하는 것이 정확성 측면에서 더 높다는 뜻이다. 또한, Jog, Sit, Stand, Walk의 Segment들은 Segment 크기를 바꾸어도 각각에 해당하는 활동을 94% 이상 예측한다. 앞선 4가지 활동들과 달리 Downstairs와 Upstairs는 Segment 크기가 변화하여도 75%보다 낮게 활동을 예측하는 것을 볼 수 있다. 특히 Segment 크기가 200일 때 Downstairs와 Upstairs Segment 예측값은 각각 56.3%, 59.6%로 50%대의 활동인식을 했다.

[표 5]는 [표 4] 실험의 LSTM 층만 2개에서 4개로 증가시키고 나머지 조건은 동일하게 진행한 결과다. Downstairs를 보면 Segment 크기가 50에서 100으로 증가 할 때 예측확률이 68.2%에서 38.5%로 약 29.7% 감소했다. 그 후 Segment 크기가 200일 때 54.1%로 또다시 증가했는데, [표 4]와 달리 [표 6]의 Downstairs에서도 비슷한 결과를 보인다. 이는 LSTM 층 개수 증가가 본 연구의 활동인식에 큰 영향을 미치지 못한다. LSTM 층을 깊게 쌓아 복잡한 문제를 해결할 수도 있지만, 학습 층(LSTM 층)의 증가가 반드시 예측모델의 성능 증가로는 이어지지 않는 것을 보여준다.

[표 4], [표 5], [표 6] 결과를 보면 Jog, Sit, Stand, Walk 4가지 활동은 Segment 크기와 LSTM 층의 개수가 변화하여도 활동인식을 90% 이상 하고 있다. 그리고 Segment 크기가 각각 50, 100, 200일 때 4가지 행동을 맞게 예측하는 확률값의 큰 변화는 없었다. 이는 측정시간에 크게 영향받지 않는다는 의미로 2.5초의 측정시간 만으로도 위의 4가지 활동의 각각 Segment에 활동 별 특징이 포함된다.

#### IV. 결론 및 향후연구

본 연구에서는 측정시간 변화와 LSTM 모델구성요소 파라미터 최적화로 활동유형의 인식을 개선 과정을 진행하였다. [표 4], [표 5], [표 6] 결과에서 Downstairs와 Upstairs를 제외한 4가지 활동은 각각의 활동을 측정 시간, 모델 구조변화와 상관없이 90% 이상 높은 예측을 하고 있다. 그리고 측정시간을 2.5초 동안만 진행해도 4가지 활동 별 특징이 인식되었다. 또한, LSTM 층 개수의 변화와 모델의 성능은 본 연구에서 크게 상관이 없는 것을 관찰했다.

위의 4가지 활동 외의 Downstairs와 Upstairs는 두 활동의 Segment 들이 각각에 해당하는 활동을 Segment 크기, LSTM 층의 개수와 상관없이 75% 미만으로 저조하게 예측하였다. 이는 본 연구에서 사용한 가속도센서 하나로 Downstairs, Upstairs의 두 동작을 구분하기 힘들다는 것을 보여주고 있다. 따라서 정밀한 측정과 행동별 특징을 효과적으로 감지하기 위해 다양한 측정 장치와 추후 연구에 필요하다. 추가 적으로 데이터 만 가지고 연구를 진행하는 것이 아닌 해당 분야 전문가들의 의견을 추가해서 모델링 하는 과정 또한 요구된다.

#### Acknowledgement

본 연구는 산업통상자원부/KEIT 지식서비스산업핵심기술개발사업 “지능정보기술 기반의 제조혁신 및 최적운영 지원 시스템의 지식베이스 프레임워크 설계 개발[20000544]” 과제의 일환으로 수행되었음.

#### References

- [1] Abdulmajid Murad and Jae-Young Pyun, “Deep Recurrent Neural Networks for Human Activity Recognition”, 2017
- [2] Yuwen Chen, Kunhua Zhong, Ju Zhang, Qilong Sun and Xueliang Zhao, “LSTM Networks for Mobile Human Activity Recognition”, 2016
- [3] Ground AI. Improving Long-Horizon Forecasts with Expectation-Biased LSTM Networks Available : <https://www.groundai.com>
- [4] Sepp Hochreiter, Jurgen Schmidhulber, “Long Short-Term Memory”, 1997
- [5] Andrej Karpathy blog, The Unreasonable Effectiveness of Recurrent Neural Networks Available : [karpathy.github.io/2015/05/21/rnn-effectiveness/](http://karpathy.github.io/2015/05/21/rnn-effectiveness/)
- [6] WISDM : Wireless Sensor Data Mining Lab. WISDM’s activity prediction dataset v1.1 Available : <http://www.cis.fordham.edu/wisdm/dataset.php#actitracker>