

Human Action Recognition based on LSTM Model using Smartphone Sensor

Yull Kyu Han
Dept. of Electronic Engineering
Tongmyong University
Busan, Korea
tu13120188@daum.net

Young Bok Choi
Dept. of Electronic Engineering
Tongmyong University
Busan, Korea
ybchoi@tu.ac.kr

Abstract— We propose a deep learning model for human action recognition in order to quickly detect occurrence of disasters such as fire and terrorism. Using the acceleration and gyroscope sensors built in the smartphone, four kinds of data on human behavior were obtained and human behavior was classified through the LSTM deep learning model. As an experiment, it was confirmed that the LSTM model can be classified 95.47% accurately.

Keywords— human action recognition; disasters detection; deep learning; LSTM

I. INTRODUCTION

Every year, many casualties are caused by various sudden disasters. In the past, natural disasters such as earthquakes, tsunamis, floods, and heavy snowfalls frequently occurred, but there are a lot of man-made disasters such as fire, terrorism and arson caused by the present. In order to evacuate quickly, it is important to recognize human behavior and judge it because it is necessary to quickly find out whether a disaster occurs.

Methods for recognizing human behavior are broadly divided into a vision based method and a sensor based method. The vision-based behavior recognition recognizes three-dimensional information of the human body by using the video image as it is, or by applying a depth sensor to the image of the camera. And it performs action recognition by using information on the body part such as the position of the joint of the human body and the joint angle [1]. The sensor-based method has a marker-based method using an optical motion sensor with a plurality of cameras and markers attached to a human body and a method using a general-purpose sensor such as a smartphone or a smart watch. Since a general wearable device exists at a certain position such as a wrist, a foot, and an eye, data of a sensor for recognizing an action can have a certain pattern and it is easy to recognize. However, smartphones can be stored in various places such as hands, bags, necks, etc. according to the individual's tendency, so that patterns of sensor data can be displayed diversely, which makes recognition difficult. Therefore, researches are being made to execute preprocessing and develop analysis algorithms to improve human recognition accuracy by using sensors of smartphones [2,3,4,5].

In this paper, in order to quickly detect the position as to whether or not a disaster occurs when a disaster such as fire or terrorism occurs in the public place, using the data of the sensor of the smartphone, the characteristic of the continuous time, we propose an algorithm for analyzing human behavior by constructing a deep learning based deep neural network. It collects human behavior data in order of time by acceleration sensor and gyroscope sensor among sensors mounted on worldwide universal smartphones and it uses RNN

(Recurrent Neural Network) and LSTM (Long Short-Term Memory) model to recognize human behavior.

II. DEEP LEARNING

A. RNN

RNN(Recurrent Neural Network) is a machine learning method that can dynamically learn data related to artificial neural networks with the concept of time series added to Neural Network. The existing DNN (Deep Neural Network) parameters are independent for each layer, but the RNN shares the hidden layer so that the computed information of the hidden layer is input again as input data to learn, classify, or predict sequential data. RNN is used for analysis and prediction of time series data such as language translation, speech recognition, texture recognition, etc.[6].

Figure 1 shows the network structure of the RNN. First, the Input data at t time is calculated by the Hidden layer, and the result is output. The output data is input as input data of Hidden layer at $t + 1$ time together with Input data of $t + 1$ time. While RNN repeats circulation in such a form, past data is reflected in the next calculation [7].

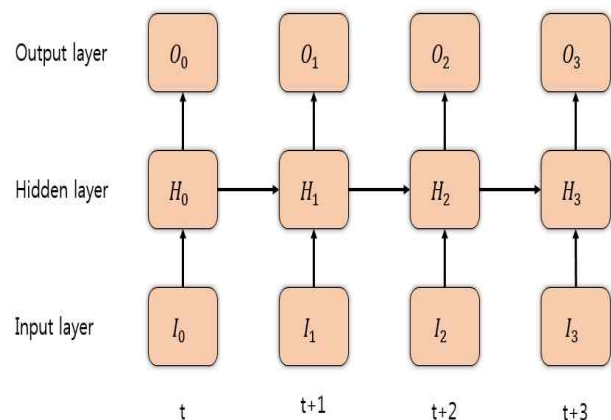


Figure 1 RNN Network Structure

However, the RNN has a long term dependency problem when the data required for the computation of the hidden layer is old. The long term dependency problem is that it is difficult to incorporate this data into the computation when the data with the key information needed for the current computation occurs a long time ago. In order to solve this problem, a LSTM (Long Short Term Memory) has been proposed.

B. LSTM

LSTM consists of three gates to solve the long-term dependency problem of RNN. Figure 2 shows the structure of LSTM. Each gate controls input, storage, and output of data to the input node on a time basis. The operation of each gate can be expressed as the product of each weight and the input value as in the basic neural network algorithm, and the information is read from the cell, input node, and the output values of other memory blocks, and the gates are activated with the sigmoid function [8].

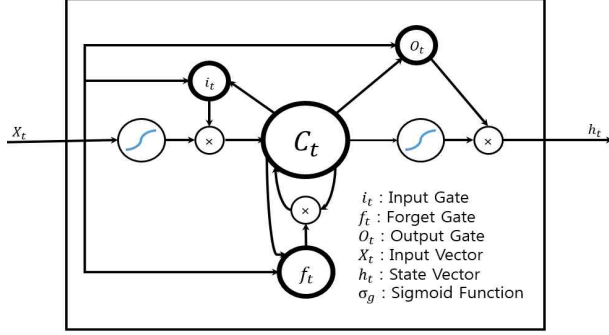


Figure 2 LSTM Structure

III. SYSTEM CONFIGURATION AND MODEL

A. System Configuration

The proposed system configuration is shown in Figure 3. The input is the cumulative data of the acceleration sensor and the gyroscope sensor of the smartphone. Cumulative data is entered into the LSTM model at consecutive time intervals. After learning with the multistage LSTM model, we classify the four behaviors of stop, walking, running, and running helter-skelter. The experiment is to learn the three data including stop, walk and running and four data including running helter-skelter in LSTM model and compare the accuracy of behavior recognition.

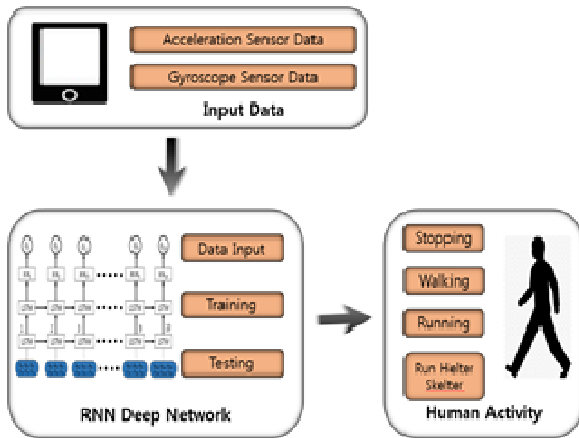


Figure 3 System Configuration

B. LSTM Model

Figure 4 shows the structure of the proposed LSTM neural network. The input layer consists of three pieces of

acceleration data, three pieces of gyro data, and a total of six pieces of input data. We input ten input data and input one input data to one LSTM node. The input data is extracted 10 data in sequence from all the data in the array. Every time input is done, 10 data are extracted again starting from the input data which is the second data of the previous input. The hidden layer consists of two layers of 10 LSTM sizes, and the first hidden layer accepts one input data from one input layer input to one LSTM. The second hidden layer has 10 outputs into the output layer. The output layer consists of the NN neural network layer. In order to distinguish four human behaviors, four outputs ultimately determine the behavior.

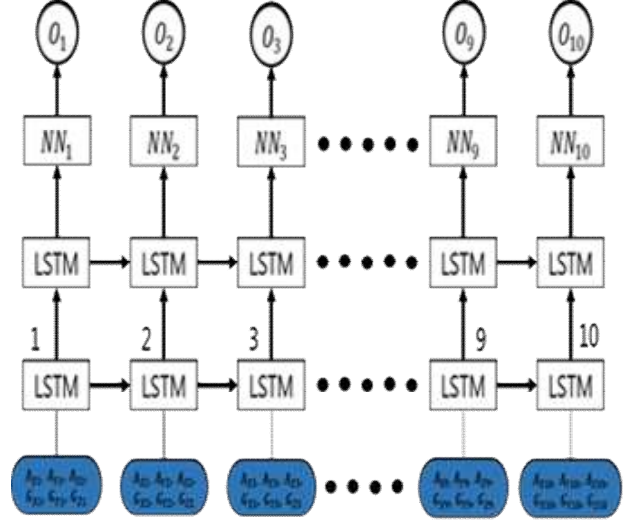


Figure 4 Deep Learning Model with NLSTM

IV. EXPERIMENTS AND RESULTS

A. Experiments

In experiments, three data including stop, walking, and running data and four data including running helter-skelter were learned as learning data in LSTM model and tested as test data. The deep Learning Model for the recognition of three behaviors is implemented by changing the output layer nodes of LSTM model described above to three. The learning data and the test data are in the case of four actions of stop, walking, running, running helter-skelter, 8000 pieces and 2000 pieces, respectively.

B. Results

Figure 5 shows the learning states of the LSTM model using the learning data on the four behaviors of stop, walking, running and running helter-skelter, and the accuracy of the predicted values using the test data. Unlike the 3 behavior learning methods, 8000 learning data were used to learn 10,000 times. As a result of testing with 2000 test data applied to the learned model, the LSTM model had an accuracy of 95.4717%. As the number of actions to be classified increased, the accuracy of the NN model fell sharply, but it was able to confirm that the accuracy of the LSTM model did not fluctuate greatly.

Step : 1000	Cost : 0.018021
Step : 2000	Cost : 0.0061564
Step : 3000	Cost : 0.0021262
Step : 4000	Cost : 0.0043949
Step : 5000	Cost : 0.00036308
Step : 6000	Cost : 0.00018102
Step : 7000	Cost : 9.8753e-05
Step : 8000	Cost : 5.6682e-05
Step : 9000	Cost : 3.3398e-05
Step : 10000	Cost : 1.995e-05
Test_data_Accuracy : 95.4717%	

Figure 5 Learning and Accuracy of Four Behaviors (LSTM)

V. CONCLUSION

In this paper, we have proposed a human behavior recognition system to quickly detect the occurrence of disasters such as fire and terrorism in public places and the location of people. We collected four behavioral data on human behavior and classified human behavior through the LSTM deep learning model. Experimental results show that the LSTM model is 95.47%.

In the future, if we improve the learning model and experiment with large amounts of experimental data, we expect to improve the accuracy.

ACKNOWLEDGMENT

“ This research was supported by Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education(NRF-2018R1D1A1B07048080)”

REFERENCES

- [1] J. Wang, Z. Liu, Y. Wu, and J. Yuan, “Learning Actionlet Ensemble for 3D HumanAction 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 Smartphones,” *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. Wen and Z. Wang, “Sensor-based Adaptive Activity Recognition with Dynamically Available Sensors,” *Neurocomputing*, Vol.218, pp. 307-317, 2016. J.I. Choi and H.S. Yong, “Activity Data
- [5] Hyunjin Lee, “Hierarchical Deep Belief Network for Activity Recognition Using Smartphone Sensor”, *Journal of Korea Multimedia Society* Vol. 20, No. 8, pp. 1421-1429, August 2017
- [6] Jong-young Park, “Estimation of Electrical Loads Patterns by Usage in the Urban Railway Station by RNN”, *The Transactions of the Korean Institute of Electrical Engineers* Vol.67, No.11, pp.1536~1541, November 2018
- [7] Joong Soo Han, Eul Gyu Im, “Implementation of Individual Gait Recognition using RNN”, *KIISE Transactions on Computing Practices*, Vol. 24, No. 7, pp. 358-362, July 2018
- [8] Myeong-Chun Lee, Sung-Bae Cho, “Mobile Activity Recognition using Bidirectional Recurrent Neural Network with Long Short-Term Memory”, *Korea Information Science Society*, Vol. 38, No. 2B, pp. 263-266, November 2011