2021 The 21st International Conference on Control, Automation and Systems (ICCAS 2021)
Ramada Plaza Hotel, Jeju, Korea, Oct. 12~15, 2021

# Prediction of pedestrian trajectory based on long short-term memory of data

Tomoya Ono[1] and Takashi Kanamaru[2*]

[1] Major of Mechanical Engineering, Kogakuin University,
Tokyo 192-0015, Japan (am20015@ns.kogakuin.ac.jp)
[2] Department of Mechanical Science and Engineering, Kogakuin University,
Tokyo 192-0015, Japan (kanamaru@cc.kogakuin.ac.jp) * Corresponding author

**Abstract:** Recently, development of navigation robots and autonomous cars are rapidly progressing. When such robots become popular in our daily life, their collisions with humans should be avoided for safety. For that purpose, we predict pedestrian trajectories with LSTM (long short-term memory) networks and conventional neural networks, and we compare their results. In order to predict sequential data, we use the following two methods: (I) predicting $n$ steps of data with $n$ models, and (II) predicting $n$ steps of data with a model by applying one-step prediction several times. By examining these two methods, it was found that the performances of the method I with the LSTM network and the conventional neural network are comparable, and the performance of the method II with the LSTM network is significantly better than that with the conventional neural network.

(Proceedings of 2021 21st International Conference on Control, Automation and Systems (ICCAS2021), pp.1676-1679.)
**Keywords:** machine learning, neural networks, long short-term memory, LSTM, pedestrian trajectory

## 1. INTRODUCTION

Recently, development of navigation robots and autonomous cars are rapidly progressing. When realizing such robots, safety for humans is highly important. For that purpose, robots are often operated with slow speed. If robots can avoid collisions with humans appropriately, such limitations for robots would be removed.

Therefore, in this study, we model the pedestrian trajectory to avoid collisions of robots with humans.

## 2. RELATED STUDIES

In order to predict the motion of pedestrians, Helbing et al. used an equation of motion in which a pedestrian is modeled as a point mass and other pedestrians and effects of obstacles are described as forces [1]:

$$m_i \frac{dv_i}{dt} = m_i \frac{v_i^0(t)e_i^0(t)-v_i(t)}{\tau_i} + \sum_{j(\neq i)} f_{ij} + \sum_W f_{iW}. \quad (1)$$

Although this model can incorporate various effects for pedestrians, it would be difficult to determine various parameters of the model.

On the other hand, we can also obtain a model of pedestrian motion based on the machine learning using only data of pedestrian motions. For example, Sakata et al. trained a Seq2Seq model that predicts the future sequential data $p_{t+1}, p_{t+2} \cdots p_{t+14}, p_{t+15}$ from the past sequential data $x_{t-14}, x_{t-13} \cdots x_{t-1}, x_t$ [2]. Alahi et al. proposed Social LSTM that is an LSTM model with a social pooling layer that can model human-human interactions, and they predicted the pedestrian trajectories [3].

Models can be trained with the machine learning when sufficient amount of data are available. Recently, it has become easy to treat large amount of data because of the improvement of computer performances and the development of software libraries. Therefore, we use a model with the machine learning in this study.

In order to predict the future sequential data $p_{t+1}, p_{t+2} \cdots p_{t+14}, p_{t+n}$ from the past sequential data $x_{t-m+1}, x_{t-m+2} \cdots x_{t-1}, x_t$, Sakata et al. [2] obtained $n$ steps of future data at once using the Seq2Seq, and Alahi et al. [3] repeated one-step predictions by adding $p_{t+1}$ to the past data. In this study, we examine the effective method for predicting $n$ steps of future data.

## 3. METHODS

When applying the prediction of pedestrian trajectory to realistic mobile robots, predicting multiple steps of data would be useful. Therefore, we predict the future sequential data $p_{t+1}, p_{t+2} \cdots p_{t+14}, p_{t+n}$ with $n$ steps from the past sequential data $x_{t-m+1}, x_{t-m+2} \cdots x_{t-1}, x_t$ with $m$ steps where $m = 15$ and $n = 10$.

For that purpose, we investigated the following two methods.

I. Predicting $n$ steps of data with $n$ models.
II. Predicting $n$ steps of data with a model by applying one-step prediction several times.

As shown in Fig. 1, with the method I, $n$ models are trained by the machine learning, and the $i$th model predicts the $i$ steps-ahead data $p_{t+i}$ $(1 \leq i \leq n)$.

On the other hand, with the method II, only one model is trained by the machine learning that predict one step-ahead data $p_{t+1}$ as shown in Fig. 2. By repeating one step-ahead prediction $n$ times, we can obtain $p_{t+n}$.

In the following sections, we compare these two methods.

**Model 1** predictions

$x_{t-m+1}, x_{t-m+2} \cdots x_{t-1}, x_t \Rightarrow p_{t+1}$

**Model 2**

$x_{t-m+1}, x_{t-m+2} \cdots x_{t-1}, x_t \Rightarrow p_{t+2}$

⋮

**Model n-1**

$x_{t-m+1}, x_{t-m+2} \cdots x_{t-1}, x_t \Rightarrow p_{t+n-1}$

**Model n**

$x_{t-m+1}, x_{t-m+2} \cdots x_{t-1}, x_t \Rightarrow p_{t+n}$

Fig. 1. Method I. Predicting $n$ steps with $n$ models

**Model** predictions

$x_{t-m+1}, x_{t-m+2} \cdots x_{t-1}, x_t \Rightarrow p_{t+1}$

$x_{t-m+2}, x_{t-m+3} \cdots x_t, p_{t+1} \Rightarrow p_{t+2}$

$x_{t-m+3}, x_{t-m+4} \cdots p_{t+1}, p_{t+2} \Rightarrow p_{t+3}$

⋮

$x_{t-m+n}, x_{t-m+n+1} \cdots p_{t+n-2}, p_{t+n-1} \Rightarrow p_{t+n}$

Fig. 2. Method II. Predicting $n$ steps with a model.

## 4. EXPERIMENTS

### 4.1 Network

In this section, we compare the two methods defined in section 3.

We treat pedestrian trajectories represented by two dimensional coordinates $(x_t, y_t)$. We define the input data as $(x_{t-m+1}, y_{t-m+1}, x_{t-m+2}, y_{t-m+2} \cdots x_t, y_t)$ where $m = 15$, and we predict the future sequential data written as $(px_{t+1}, py_{t+1}, px_{t+2}, py_{t+2} \cdots px_{t+n}, py_{t+n})$ where $n = 10$.

As models of the machine learning, we use a long short-term memory (LSTM) network and a conventional multi-layer neural network (NN). The LSTM network is a kind of recurrent neural network (RNN) and it is known that the LSTM network is suitable for learning sequential data. For the NN, we set the number of hidden layers as 1, which is composed of 300 neurons.

In both cases, the networks are trained to minimize mean squared error (MSE).

### 4.2 Dataset

As a dataset, we use the ATC pedestrian tracking dataset [4] of pedestrian trajectories of general shoppers in the ATC shopping center in Osaka, which includes 3,758,348 trajectories obtained from various sensors at 10−40 [Hz].

In this study, we chose only the trajectories confined in an area with 10 [m] × 10 [m], and we use 200 trajectories for training, and 20 trajectories for test.

### 4.3 Method I

MSE of each prediction step with the method I for the LSTM network and the NN are shown in Fig.3. The mean and the standard deviation of MSE of 20 test data are shown.

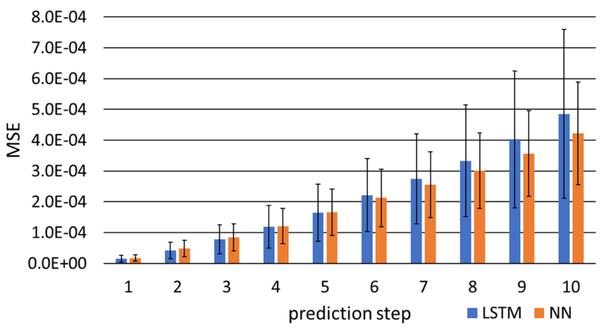


Fig. 3. MSE of each prediction step with the method I for the LSTM network and the NN.

In Fig. 3, the significant difference ( $p < 0.05$ ) between the results of the LSTM network and the NN was not observed.

Some examples of the predicted trajectories are shown in Fig. 4. The data with blue line show the original data, and the data with orange line show the predicted data $p_{t+10}$.

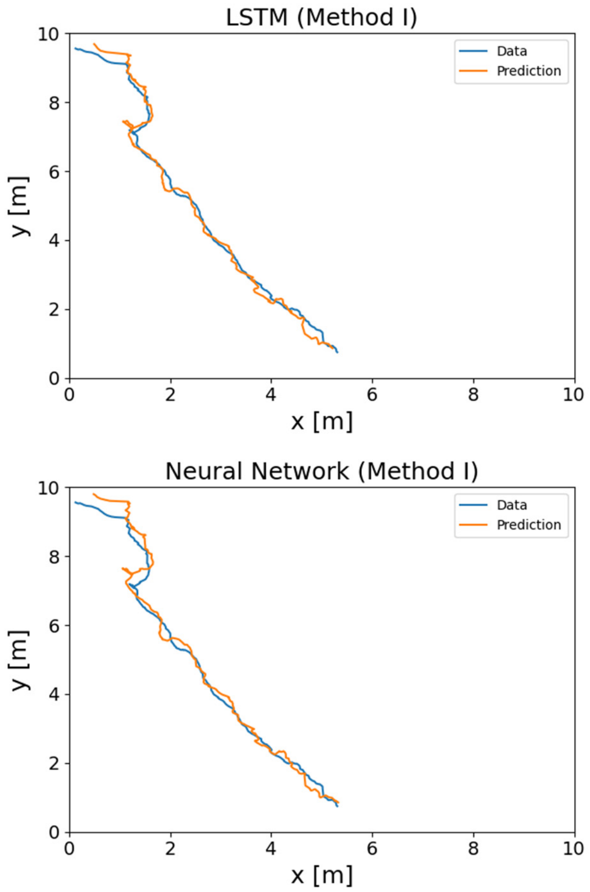


Fig. 4. Method I. Predicting $n$ steps with $n$ models.

### 4.4 Method II

MSE of each prediction step with the method II for the LSTM network and the NN are shown in Fig.5. The mean and the standard deviation of MSE of 20 test data are shown.

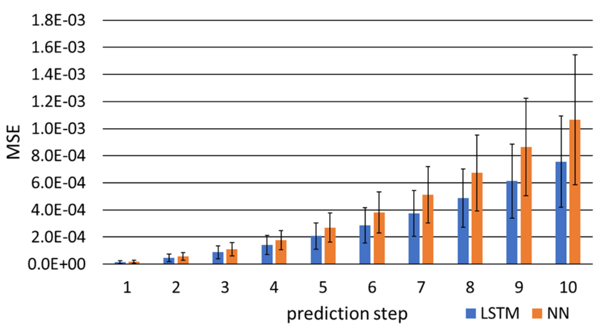


Fig. 5. MSE of each prediction step with the method II for the LSTM network and the NN.

In Fig. 5, the significant difference ($p < 0.05$) between the results of the LSTM network and the NN was observed.

Some examples of the predicted trajectories are shown in Fig. 6. The data with blue line show original data, and the data with orange line show the predicted data $p_{t+10}$.

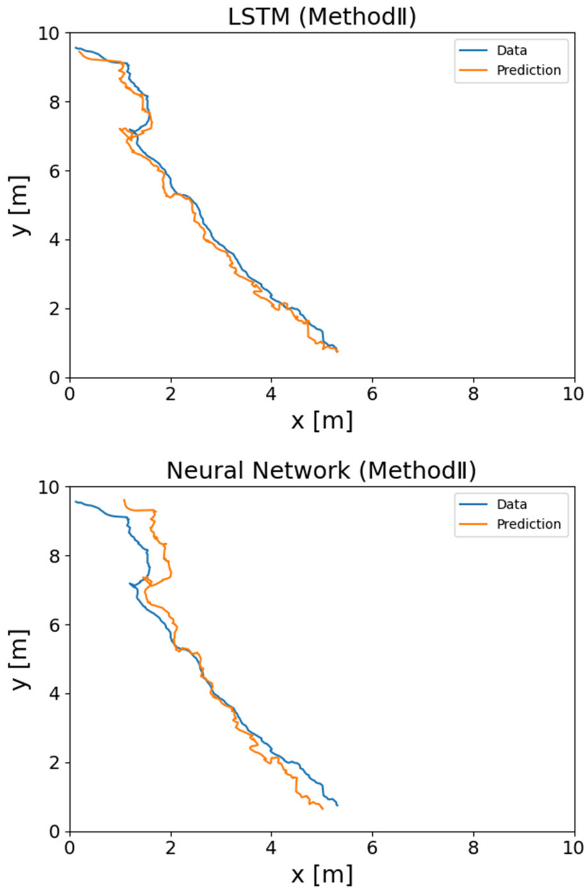


Fig. 6. Method I. Predicting $n$ steps with $n$ models.

### 4.5 Comparison

By comparing the results of the method I and II, it is observed that the predictions with the method I give better results than those with the method II as shown in Figs. 3 and 5. This result would not be surprising because the method I can use multiple steps-ahead data $x_{t+i}$ $(1 \leq i \leq n)$ for training.

By comparing the results of the LSTM network and the NN, it is observed that there was no significant difference between the results of the LSTM network and the NN with the method I as shown in Fig. 3.

On the other hand, with the method II, the LSTM network shows better prediction than that of the NN as shown in Fig. 5. It would be because that the LSTM network is suitable for learning sequential data.

With the method I, the trained data were not sequential, i.e., the multiple steps-ahead data $x_{t+i}$ $(1 \leq i \leq n)$ were trained with the $i$th LSTM network; therefore, each

LSTM network could not show better results than that of the NN.

## 5. CONCLUSION

In this study, we investigated two methods for predicting pedestrian trajectories. In order to predict the sequential data, we used the following two methods: (I) predicting $n$ steps of data with $n$ models, and (II) predicting $n$ steps of data with a model by applying one-step prediction several times.

It was found that the method I gives better predictions than those with the method II. However, in order to use the method I, we have to train $n$ models using multiple steps-ahead data $x_{t+i}$ $(1 \leq i \leq n)$.

When solving realistic problems, it would not be possible to train arbitrarily large number of models, and predictions exceeding the range of training would often be required. In such cases, the method II would be important.

When using the method II, it was found that the LSTM networks give better predictions than those of the conventional neural networks. By improving the method II with the LSTM network, this method would become applicable to predictions for general purposes.

In this study, we did not use interactions among pedestrians and effects from obstacles. By incorporating such effects into our model, further improvement would be possible.

## REFERENCES

[1] D. Helbing, and P. Molnar, "Social Force Model for Pedestrian Dynamics", Physical Review E, Vol. 51, No. 5, pp. 4282-4286, 1995.

[2] N. Sakata, Y. Kinoshita and Y. Kato, "Predicting a Pedestrian Trajectory Using Seq2Seq for Mobil e Robot Navigation", The 44nd Annual Confer ence of the IEEE Industrial Electronics Society, pp. 4300-4305, 2008.

[3] A. Alahi, K. Goel, V. Ramanathan, A. Robicquet, Li Fei-Fei, S. Savarese, "Social LSTM: Human Trajectory Prediction in Crowded Spaces", Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 961-971, 2016.

[4] D. Brscic, Y. Kanda, T. Ikeda and T. Miyashita, "Person Position and Body Direction Tracking in Large Public Spaces Using 3D Range Sensors," IEEE Transactions on Human-Machine Systems, Vol. 43, No. 6, pp. 522-534, 2013.