Pedestrian Destination Prediction Using a Pretrained Predictor Model for a Voice Guidance Robot

Yuka Kato and Asami Ohta[1]

I. INTRODUCTION

In recent years, there has been increasing interest in communication robots that can communicate with a person through sound generation, head or arm movement, etc. A variety of services such as voice guidance are expected for such robots. However, providing appropriate services to persons is quite difficult because there are various environments where robots are installed and various conditions where robot users are located. A state estimation of robot users is required.

From these backgrounds, we have been studying methodologies that estimate the situations of robot users from sensing results with a laser range scanner (LRS) attached to a communication robot. In particular, we focused on guidance robots installed at the reception desks of public facilities. For example, we proposed a method that detected a person approach and changed the content of speech [1], and a classification method for age groups (elderly and young people) using the gait [2]. In these cases, the trajectories of persons were varied, and the destinations were different depending on their purpose. Therefore, the manner of voice guidance should also depend on the purpose. However, the previous methods only estimated the current situation, and failed to consider suitable guidance for the purposes of users.

To solve these problems, in this paper, we attempt to predict the destination by using the trajectory of a pedestrian, and use the prediction result for the voice guidance of the communication robot. As a preliminary study, for the reception desk as shown in Fig. 1, we designed a method to predict the walking direction (”office side” or ”reception side”) when the pedestrian came within the sensing range of the LRS. In the study, we measured the orientation to the reception desk of the person’s body while tracking the pedestrian, and used the data for the prediction. The benefit of this approach is that the branching direction can be easily predicted within a limited time from the data in a limited measurement range by using a single sensor attached to the robot. However, there are certain drawbacks associated with the use of a single data of body orientation, such as the insufficient accuracy. The aim of this paper is to enhance this method and to improve the accuracy of prediction by using instantaneous body orientation data and the trajectory data of a pedestrian. Specifically, we train a predictor model using massive trajectory data by a machine learning algorithm, and predict the branching direction by using the predictor model.

[1]Yuka Kato and Asami Ohta are with the Division of Mathematical Sciences, Tokyo Woman’s Christian University, 2-6-1 Zempukuji Suginami-ku, Tokyo 1678585, Japan. yuka@lab.twcu.ac.jp

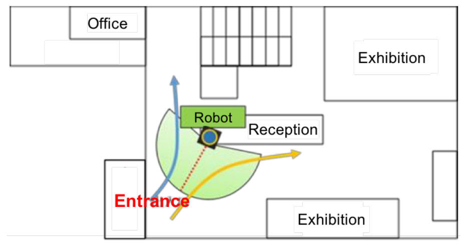

Fig. 1. Layout of target reception desk. LRS is placed at black square position in middle of figure. Pedestrians enter building from entrance and head to LRS. Branching directions are predicted by using sensing data.

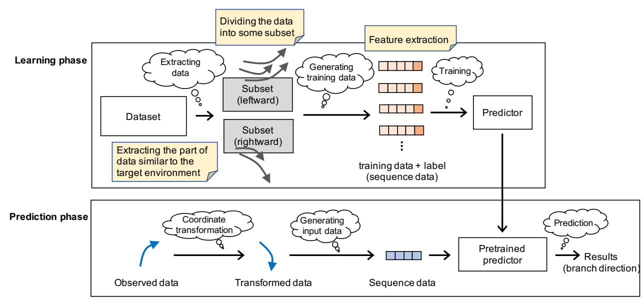

Fig. 2. Outline of the proposed method. The method predicts a branch direction of a pedestrian with a pretrained predictor model.

At this time, massive trajectory data is necessary to train the predictor model; however, collecting a sufficient amount of data is difficult in general public spaces and may cause difficulties in model development. In this paper, we build the following hypothesis: trajectories of pedestrians are generally similar under similar conditions, and we can generate the predictor model without depending on the environment. Based on this hypothesis, the proposed method trains the predictor model using an open dataset including massive pedestrian trajectory data, and predicts the branch direction in the target environment by using the model. The outline of the method is shown in Fig. 2. As for features, it was considered that utilizing velocity vectors would effectively predict the branch direction, and the time sequence of the vector would be meaningful as well. Therefore, we use the sequence of the velocity vector for each time step derived from measurement data within the area for judging. The contributions of this paper include:

- Introduction and solution of a method to predict the branch direction of a pedestrian with a pretrained predictor model using an open dataset.
- Demonstration that the sequence of the velocity vector is effective as a feature of the model.

II. RELATED WORKS

So far, a considerable amount of literature has been published on destination prediction methods. These studies are divided into two types. One type formulates pedestrian dynamics as a mathematical model and estimates the trajectory according to a pedestrian model. The other type estimates the pedestrian trajectory by using massive observation data.

As for the former type, one well-known early study is Helbing's Social Force Model (SFM) [3]. The SFM defined psychological and physical repulsion between a person and a person, and a person and a wall, and predicted the traveling direction by the repulsion. Similarly, a pedestrian model in a human-robot coexisting environment [4] were proposed. The advantage of these studies is a lack of necessity for massive observation data and the easy utilization of the method. However, to use the models, we must set a large number of parameter values. Thus, it becomes difficult to adjust the model for an accurate prediction.

In contrast to the studies described above, others have taken a different approach by focusing on collecting massive data. To date, a number of studies attempted to generate pedestrian models using observation data, including a method that generated the model based on massive trajectories and used it for mobile robot path planning [5]. In recent years, many authors began to apply machine learning algorithms, especially using natural language processing techniques, to such models. For example, a method to predict the human trajectory in crowded spaces using long short-term memory (LSTM) [6] were proposed. These methods can accurately predict pedestrian trajectories in a short time when generating an appropriate predictor model. However, they require collecting massive data for each environment beforehand.

In this paper, we use a pretrained predictor generated in an environment different from that of the target. Consequently, we overcome the problem that collecting a sufficient amount of data is difficult in public spaces.

III. ROBOT SYSTEM

A. Overview of the system

The architecture of the target robot system is shown in Fig 3. The system consists of a communication robot, an LRS attached to the robot (the scan range is 270° and the maximum detection distance is 10,000 m), and a laptop PC connected to the LRS. In this system, when a pedestrian enters the measurement range (4000 mm), the LRS measures the distance and angle to the pedestrian. At that time, the data is processed on the PC, and robot control commands for guidance are generated by the PC. Then, control commands are sent to the robot by socket communication.

The system is installed at the reception desk of a museum, and the robot guides guests with information about the museum's facilities and exhibitions. In this research, we predict the walking direction of a guest to the "office side" or "reception side," and change the guidance content according to the direction. Here, we define the reception side as "rightward" and the office side as "leftward".

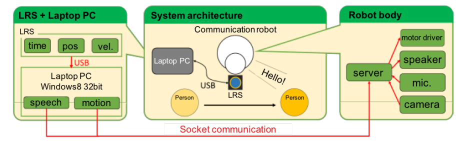

Fig. 3. Overall configuration of target robot system. LRS is attached to robot, but sensing data with LRS is processed on laptop PC connected to LRS. Communication robot runs according to processing results.

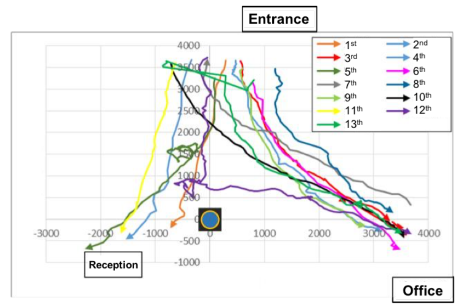

Fig. 4. Result of preliminary experiment to investigate trajectory trend. Trajectories for 13 pedestrians are illustrated by arrows.

B. Preliminary experiment to investigate the trajectory trend

In order to investigate the trajectory trend in the target environment, we conducted a preliminary experiment. The results are shown in Fig. 4. We obtained trajectory data for 13 pedestrians. As for the axis of the coordinates, the origin is the position where the LRS is installed, the y axis is the direction of the central axis of the LRS, and the x axis is the horizontal direction. The unit is mm. As shown in Fig. 4, we found there is a possibility that the trajectory tendency to the reception side and the office side can be estimated from a geometric shape of each trajectory.

C. Response timing of robot

In the target environment, the system must finish judging the direction within the period such that the speech of the robot can be generated before the pedestrian reaches reception. According to a previous work, the communication time to transmit a command from the PC to the robot was 40 ms, and the operation time of the robot was 985 ms. Here, the speech is generated after the motion. In addition, it is known that the average normal walking speed of a Japanese adult male is 1400 mm/s. These facts indicate that the distance from the robot to the judging area should be more than $(40+985) \times 1400/1000 = 1435$ mm. In this paper, since the measurement range of the LRS is 4000 mm, we determine that the system judges the branching direction in the range of $2000 \leq y \leq 3500$, and generates the motion and speech of the robot in a range of $0 \leq y < 2000$.

IV. PROPOSED METHOD

A. Outline of the method

In this paper, we train a predictor model using an open dataset, and predict the direction in a real environment using the pretrained model. The procedure is divided into two phases: the learning phase and the prediction phase.

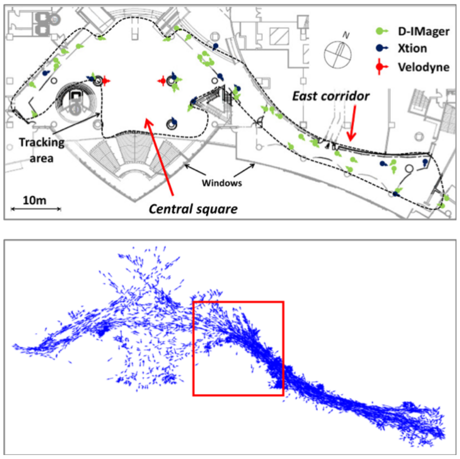

Fig. 5. (Upper figure) Shape of shopping mall where data were measured, referred from [7]. This is a combination of a plaza where people are staying and a passage where people are walking. (Lower figure) Result of visualization of part of trajectory data. In the proposed method, we extract data for a number of targeted pedestrians from data in red frame.

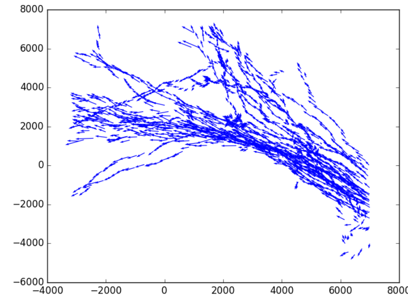

Fig. 6. Example of extracting 30 trajectories branching to right and 30 trajectories branching to left around branch point.

B. Learning phase

1) Dataset: To train a predictor model, we use the ATC pedestrian tracking dataset [7] as normal walking data in public facilities. This is a dataset of movement trajectories of 3,758,348 pedestrians (the average is 40,851 per day, including duplication). This is data for 92 days for about one year (from 9:40 to 20:20 on every Wednesday and Sunday from Oct. 2012 to Nov. 2013). Human positions were continuously measured at $10 \sim 40$ Hz with multiple 3D range image sensors. The shape of the shopping mall is shown in the upper part of Fig. 5. The data is stored in CSV format, and the following items are contained in each record: time [ms]; person ID; position x, y, z [mm]; velocity [mm/s]; angle of motion [rad]; facing angle [rad]. Here, we denote x, y, z at time t as x_t, y_t, z_t, the velocity as v_t, and the angle of motion as θ_t.

In order to determine the movement image, we picked up some tracking data and visualized them. The lower part of Fig. 5 shows that. Here, we extract the trajectory of a certain pedestrian for N persons and use this to train the model.

2) Generating training data: The training data are generated by extracting the part of data similar to the target environment from the dataset. Specifically, we regard the area in the red frame in the lower part of Fig. 5 as a branch point. Let T be a set of all trajectories of all pedestrians passing through this area, $T^r \subset T$ a set of trajectories that branch rightward, and $T^l \subset T$ a set of them that branch leftward.

First, we randomly select $N/2$ target trajectories from elements of T^r and T^l. Fig. 6 shows an example of visualizing the selected trajectory data ($N = 60$). Here, we also indicate the coordinate axis in the dataset. Next, we generate features for training as sequence data for each trajectory. From the preliminary experiment, we assume that the branching direction can be expressed by constructing each trajectory data as a set of velocity vectors $\boldsymbol{a}_t = (v_t, \theta_t)$. In this case, we discretize the trajectory data by considering only the arrangement order and without considering the time interval. In order to give meaning to the order of $\boldsymbol{a}_t$, we introduce a new parameter i to express the order. Here, assuming that time t at which the data assigned i is observed is $t(i)$, $t(i) < t(i+1)$ always holds. As a result, a certain movement trajectory $\mathcal{T} \in T$ is defined as

$$\mathcal{T} \; : \; = \; \boldsymbol{a}_0, \; \ldots, \; \boldsymbol{a}_i, \; \boldsymbol{a}_{i+1}, \; \ldots \tag{1}$$

as an ordered pair of $\boldsymbol{a}_i$.

Since the area for judging the direction in the target environment is the part between 2000 mm and 3500 mm from the branch point, we can generate the training data by extracting $\boldsymbol{a}_i$ corresponding to this area while keeping the order relation from $\mathcal{T}$. That is, we generate a set $\mathcal{D} \subset \mathcal{T}$ consisting of $\boldsymbol{a}_k$ corresponding to k that satisfies $2000 \leq x_k \leq 3500$, and use a set D^r consisting of $\mathcal{D}$, which is a subset of $\mathcal{T} \in T^r$, and a set D^l consisting of $\mathcal{D}$, which is a subset of $\mathcal{T} \in T^l$, as the training data. At that time, $\mathcal{D}$ is constructed by arranging v_i and θ_i in chronological order, and n velocity vectors are extracted in ascending order of x_k in order to adjust the number of data among trajectories. As a result, we can express $\mathcal{D}$ as follows:

$$\mathcal{D} \; : \; = \; v_0, \; \theta_0, \; \ldots, \; v_{n-1}, \; \theta_{n-1}. \tag{2}$$

3) Generating a predictor model: In this paper, we use a Support Vector Machine (SVM) as a predictor model.

C. Prediction phase

1) Measuring data: In the prediction phase, we track the position of a pedestrian in the LRS measurement range ($y \leq 4000$) and measure the distance and angle from the LRS to the pedestrian at a frequency of 10 Hz. For tracking, we use the same method as described in [2]. Here, the position of the tracking target is expressed as a probability distribution, and the state of the tracking object is recursively estimated using a probability density function from a transition model and an observation model. We use a particle filter for the probability distribution of the position of a pedestrian.

At that time, the position is calculated from the distance and angle measured by the LRS by converting the polar coordinate format to the orthogonal one. From the result, pedestrian position data $(\hat{x}_t, \hat{y}_t)$ per unit time are obtained.

2) Generating input data: From the measurement data, we generate input data for the predictor model. At that time, sequence data of the velocity and angle are input to the model, and we have to prepare data in the same format as the data used for the model. Specifically, we transform the coordinate system of the target system shown in Fig. 4 into that of the dataset shown in Fig. 6, and adjust the sequence length of data to $2n$. As for coordinate transformation, absolute values of x and y are meaningless, so we conduct only a rotational transform. As for adjusting the length, if the sequence length is greater than $2n$, it is aborted with $2n$; and if that is less than $2n$, it compensates for missing data.

First, rotate $(\hat{x}_t, \hat{y}_t)^T$ by φ, and calculate $(x_t, y_t)^T$ corresponding to the coordinate axis of the dataset, as follows:

$$\begin{pmatrix} x_t \\ y_t \end{pmatrix} = \begin{pmatrix} \cos\varphi & -\sin\varphi \\ \sin\varphi & \cos\varphi \end{pmatrix} \begin{pmatrix} \hat{x}_t \\ \hat{y}_t \end{pmatrix}. \quad (3)$$

Here, we set $\varphi = -3\pi/4$. Then, we calculate v_t and θ_t by

$$v_t = \sqrt{{\Delta x_t}^2 + {\Delta y_t}^2}\ /\Delta t \quad (4)$$

$$\theta_t = \begin{cases} \arctan(\Delta y_t/\Delta x_t) & (\Delta x_t > 0) \\ \arctan(\Delta y_t/\Delta x_t) + \pi & (\Delta x_t < 0,\ \Delta y_t \geq 0) \\ \arctan(\Delta y_t/\Delta x_t) - \pi & (\Delta x_t < 0,\ \Delta y_t < 0) \end{cases} \quad (5)$$

where Δt is unit time, and $\Delta x_t = x_{t+1} - x_t$, $\Delta y_t = y_{t+1} - y_t$. As a result, we generate $2n$ sequence data $\mathcal{D}^*$ as follows:

$$\mathcal{D}^* := v_0,\ \theta_0,\ \ldots,\ v_t,\ \theta_t,\ \ldots,\ v_{n-1},\ \theta_{n-1}. \quad (6)$$

3) Predicting the direction: In this step, we predict the branch direction by inputting $\mathcal{D}^*$ to the predictor model.

V. EXPERIMENT

First, we verified the performance of the predictor model by using the dataset. Next, we confirmed that this model can predict the branch direction using observed data.

A. Experiment 1: Verification using dataset

1) Experimental settings: Using the ATC dataset, we conducted an evaluation experiment using the generated model. Here, we changed N to 40, 100, 200, and 400, and obtained precision, recall, and F values for the data branching to the right. We used $n = 23$. For data verification, we divided the data into 5 groups and performed a cross-validation. To generate the model, we used the svm_SVC of scikit-learn.

2) Experimental result: The experimental result is shown in Table I. From this result, we found out that predicting the branch direction is possible with an accuracy of about 80% if we obtain about 100 training data. Considering this tendency and the application of the proposed method, the classification accuracy is sufficient.

B. Experiment 2: Verification using observed data

1) Experimental settings: Using the data observed in the target environment, we evaluated the prediction result. For the observation data, we chose the data for four pedestrians in Fig. 4. In this case, as typical trajectories, we chose the 1st and 2nd persons (to the right), and 8th and 10th persons (to the left). We extracted 24 coordinates from each trajectory.

2) Experimental result: The experimental result is shown in Table II. Here, "0" indicates the branch to the right, and "1" to the left. We generated the model by changing N to 40, 100, 200, and 400. In all cases, the classification results were correct. Although the pattern of using data was limited, we were able to verify the effectiveness of prediction.

VI. CONCLUSIONS

The results of the paper are as follows:

- The method was proposed using a pretrained predictor model with a dataset by extracting the part of data similar to the target environment.
- We verified that we can predict the branch direction of a pedestrian with an accuracy of about 80%.
- We verified that prediction with sufficient accuracy is possible in the target environment if we obtain training data for about 100 pedestrians.

The subject of consideration in this paper seems to be valid as a setting for communication robots. However, system conditions vary according to the environment. More detailed studies will be necessary in order to examine the influence of such environment-dependent conditions.

REFERENCES

[1] D. Nomiyama, N. Matsuhira, M. Sano, and T. Yamaguchi, "Enhancement of interface robot using rt middleware and rsnp network protocol," in *Proc. IEEE Int. Workshop on Advanced Robotics and its Social Impacts (ARSO 2015)*, 2015, pp. 1–6.

[2] S. Sakai, S. Kimura, D. Nomiyama, T. Ikeda, N. Matsuhira, and Y. Kato, "Classification of age groups using walking data obtained from a laser range scanner," in *Proc. the 42nd Annual Conference of the IEEE Industrial Electronics Society (IECON 2016)*, 2016, pp. 5862–5867.

[3] D. Helbing and P. Molnar, "Social force model for pedestrian dynamics," *Phys. Rev. E.*, vol. 51, no. 5, p. 4282, 1995.

[4] Y. Kato, Y. Nagano, and H. Yokoyama, "A pedestrian model in human-robot coexisting environment for mobile robot navigation," in *Proc. 2017 IEEE/SICE International Symposium on System Integration (SII 2017)*, 2017.

[5] H. Noguchi, T. Yamada, T. Mori, and T. Sato, "Mobile robot path planning using human prediction model based on massive trajectories," in *Proc. IEEE International Conference on Networked Sensing Systems (INSS 2012)*, June 2012, pp. 1–7.

[6] A. Alahi, K. Goel, V. Ramanathan, A. Robicquet, L. Fei-Fei, and S. Savarese, "Social lstm: Human trajectory prediction in crowded spaces," in *Proc. 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR 2016)*, 2016, pp. 961–971.

[7] D. Brscic, T. Kanda, T. Ikeda, and T. Miyashita, "Person position and body direction tracking in large public spaces using 3d range sensors," *IEEE Trans. on Human-Machine Systems*, vol. 43, no. 6, pp. 522–534, 2013.

TABLE I

COMPARISON RESULT OF PREDICTION PERFORMANCE BY NUMBER.

Number of training data	40	100	200	400
Precision	0.75	0.81	0.86	0.82
Recall	0.63	0.78	0.80	0.82
F value	0.69	0.79	0.83	0.82

TABLE II

PREDICTION RESULT OF BRANCH DIRECTION USING OBSERVED DATA.

Number of training data	40	100	200	400
Result of 1st person (right)	0	0	0	0
Result of 2nd person (right)	0	0	0	0
Result of 8th person (left)	1	1	1	1
Result of 10th person (left)	1	1	1	1