Advance Publication by J-STAGE

日本機械学会論文集

Transactions of the JSME (in Japanese)

DOI:10.1299/transjsme.21-00125

Received date : 31 March, 2021

Accepted date : 2 June, 2021

J-STAGE Advance Publication date : 10 June, 2021

周辺環境の障害物を考慮した深層学習による歩行者の軌道予測

杉浦 尚弥^{*1}, 松田 匠未^{*2}, 黒田 洋司^{*3}

Pedestrian trajectory prediction by deep learning considering obstacles in the surrounding environment

Naoya SUGIURA^{*1}, Takumi MATSUDA^{*2} and Yoji KURODA^{*3}

^{*1} Graduate School of Science and Technology, Meiji University

1-1-1 Higashimita, Tama-ku, Kawasaki-shi, Kanagawa 214-8571, Japan

^{*2,*3} School of Science and Technology, Meiji University

1-1-1 Higashimita, Tama-ku, Kawasaki-shi, Kanagawa 214-8571, Japan

Abstract

In this paper, we propose a pedestrian trajectory prediction method for autonomous mobile robots. In many cases, there are many pedestrians in the environment in which the autonomous mobile robot runs. In such an environment, the robot needs to run safely for pedestrians. In order to avoid collisions with pedestrians and drive safely, it is important to predict future movements of pedestrians. In the conventional prediction method, the trajectory of a future pedestrian is often predicted from the position of a pedestrian in the past. However, in such cases, it is difficult to predict the movement to avoid obstacles such as walls and pillars around the pedestrian. In this study, point cloud data acquired by LiDAR is used to predict the behavior of pedestrians avoiding surrounding static obstacles. Based on the point cloud data, distances between the pedestrians and the static obstacles is calculated at each time. Then, input it into the network together with the transition of the pedestrian's position to predict the future pedestrian's position. In addition, we use attention mechanisms to model interactions between pedestrians. This makes predictions that consider not only static obstacles but also the effects of other pedestrians. The usefulness of this study is shown by performing accuracy evaluation using the dataset created in the simulation environment and the publicly available dataset.

Keywords : Trajectory prediction, Mobile robot, Static obstacle, Attention mechanism, Spatial interaction

1. 緒 言

近年、自律移動ロボットの研究が盛んに行われており、建物の警備や配達等の業務を目的としたものなど、人々が生活する実環境でロボットが活躍する場面が多く見られるようになってきている。人のいる環境で活動するロボットは、歩行者の行動を妨げないような社会性を持った行動をとることが求められる。歩行者との衝突を回避するためには、歩行者に近づきすぎないように事前に回避行動をとる必要があり、そのためには、歩行者の将来とるであろう行動を予測することが重要である。歩行者の動作は、歩行者の周囲に存在する他の歩行者などの動的障害物や、進路上に存在する木や、建造物の壁や柱などの静的障害物からの影響により複雑に変化する。周囲の状況に応じて多様に変化する歩行者の行動を予測するためには、進路上に存在する障害物や、他の歩行者との相

No.21-00125 [DOI: 10.1299/transjsme.21-00125]

^{*1} 学生員, 明治大学大学院 理工学研究科 (〒214-8571 神奈川県川崎市多摩区東三田 1-1-1)

^{*2} 明治大学 理工学部

^{*3} 正員, 明治大学 理工学部

E-mail of corresponding author: ce202036@meiji.ac.jp

相互作用をモデル化する必要がある。

歩行者の軌道を予測する手法は、オンボードカメラなどで観測した一人称視点における歩行者位置に基づいて予測を行うもの (Yagi et al., 2018) (Styles et al., 2020), 歩行者位置を鳥瞰視点 (x - y 座標系) で扱うもの (Alahi et al., 2016) (Gupta et al., 2018) の 2 つに分けることができる。多くの移動ロボットはナビゲーションをする際に、周囲の物体とのユークリッド距離を基に経路計画を行う。本手法では、予測結果を移動ロボットが経路計画をする際に使用することを想定しているため、鳥瞰視点で歩行者位置を扱う。

Helbing らは、歩行者と目的地の間に引力を、周囲の障害物や歩行者との間には斥力を定義し、歩行者の動作をモデル化することで、歩行者の動作を予測する手法を提案した (Helbing and Molar, 1995)。しかし、この手法はパラメータを決定するのに多大な労力を要し、環境に応じて再調整する必要があるため、汎化性能が低いという欠点がある。また、高梨らは、歩行者の速度と進行方向を確率的にモデル化し、歩行者が障害物を回避する際の軌跡の予測を予測する手法を提案した (高梨他, 2017)。この手法は駐車車両などの静的障害物に対する回避行動に焦点を当てており、他の歩行者などの動的障害物存在は考慮されていないため、歩行者が多数存在する環境では正確な予測を行うことが困難であると考えられる。

近年の研究では、群衆の相互作用を捉えるために、ニューラルネットワークを使用するものが多く存在する。Alahi らは、群衆の中の歩行者間の相互作用を表現するために、Social Pooling を提案した (Alahi et al., 2016)。しかし、このモデルは近くに存在する歩行者のみがお互いの動きに影響を与えることを前提としており、実際の環境において必ずしも当てはまるものではない。また、Mohamed らは、各歩行者の状態をグラフとしてモデル化し、CNN により将来の軌道を予測する手法を提案した (Mohamed et al., 2020)。上記二つの手法は、人同士の相互作用のみに焦点を当てており、周囲に存在するその他の物体が与える影響を考慮していない。多くの場合、歩行者が存在する環境には、歩行者や自動車などの動的障害物の他に、建造物の壁や、柱などの静的障害物が存在しており、これらが歩行者に与える影響も考慮する必要がある。Veličković らはグラフ構造のノードに異なる重要度を暗黙的に割り当てるために Graph Attention Network (GAT) を提案した (Veličković et al., 2017)。Huang らは歩行者の相互作用をグラフとして定式化することにより、この概念を軌道予測に適用した (Huang et al., 2019)。Sadeghian らはシーンのコンテキストを入力することにより、周囲の環境情報を組み込むことを試みた (Sadeghian et al., 2019)。しかし、本手法で用いられている俯瞰画像のようなデータは限定された領域でしか取得することができず、移動ロボットのように活動領域が多様な場合、このようなデータを利用するのは困難である。

本研究では、他の歩行者などの動的障害物だけでなく、周囲の環境に存在する静的障害物からの影響を考慮した軌道予測手法を提案する。従来の位置情報のみを用いた軌道予測手法では、周囲の障害物からの影響を考慮することができなかった。また、Social Force Model (Helbing and Molar, 1995) などのモデルベースでの予測手法では汎化性能が低く、長期的な予測が困難であった。そこで、本手法では、歩行者の過去の位置の遷移に加え、障害物マップから算出した対象歩行者と障害物の位置関係を時系列データとしてネットワークの入力することで、静的障害物からの影響を考慮した軌道予測モデルを構築する。障害物との距離情報の変化とは別に、他の歩行者との関係性がもたらす影響を学習することで、歩行者が静的障害物を回避する動作と、移動障害物を回避する動作の両方を予測することが期待される。

2. 周辺環境の時空間的变化を考慮した歩行者の軌道予測

本章では、歩行者の周辺に存在する他の歩行者などの動的障害物と、壁や柱などの静的障害物からの影響を考慮した歩行者の軌道予測について述べる。LiDAR により取得した点群から占有格子地図を作成し、歩行者と障害物の距離を算出する。算出した距離情報と、歩行者の過去の軌道を入力とし、歩行者に対する周辺環境の影響を考慮したアテンションモデルを構築する。これにより、歩行者が静的障害物から受ける影響と、他の歩行者のような移動障害物から受ける影響をモデル化し、学習を行う。学習した予測モデルを用いて、歩行者が他の歩行者との関係性や、周囲の障害物との距離の変化を基に行動を変化させることを予測することを目指す。

Fig. 1 Architecture for the proposed model. The model consists of encoder and decoder module. We use the beam model to capture the features of the environment. The model makes use of a graph attention network to consider the interactions between pedestrians. Lin represents an embedding layer, GAT represents Graph Attention layer and MLP represents a multilayer perceptron in the figure.

2.1 障害物情報の取得

本手法では各時刻において、LiDARを用いてロボットの周囲に占有格子地図を生成し、Thrunらによって提唱された仮想ビームモデル (Thrun et al., 2005) を用いて地図上の占有された格子と歩行者の距離を算出する。仮想ビームモデルとは、空間を均等な角度のビームゾーン領域に区切り、基準点から障害物までの距離を測るモデルのことである。仮想ビームモデルを式 (1) に示す。

$$\begin{cases} Z_k = \left\{ \frac{(k-1)\pi}{n} < \tan^{-1} \left(\frac{y-y_{ini}}{x-x_{ini}} \right) \leq \frac{k\pi}{n}, |x| \leq r_{max}, |y| \leq r_{max}, k = 1, 2, 3, \dots, n \right\} \\ d_k = \left\{ \min \sqrt{x_j^2 + y_j^2}, P_j \in Z_k, k = 1, 2, \dots, n \right\} \end{cases} \quad (1)$$

ここで、 Z_k は k 番目のビームゾーン領域、 n は分解能、 d_K はビームゾーン k 内の点 P_j 間の最短距離、 r_{max} はビームの最大の長さ、 x_{ini}, y_{ini} は仮想ビームモデルの発射点を意味する。本手法では、仮想ビームモデルの基準点を、観測した歩行者位置とする。取得した歩行者と障害物の距離情報は、Min プーリングにより圧縮し、正規化したものをネットワークへ入力する。式 (2) に Min プーリングを示す。

$$s_k = \frac{d_k}{r_{max}} \quad (2)$$

ここで、 k は 1D プーリングにおけるカーネルサイズ、 l はカーネルのインデックスである。

2.2 ネットワーク構造

動的なシーンでは、各歩行者は他の動的な歩行者の動きとシーン内の固定された障害物の制限に基づいてナビゲーションをする。本手法では、観測した歩行者の過去の軌道と、周囲の障害物と歩行者との位置関係を用いて将来の歩行者の軌道の予測を行う。 i 番目の歩行者の過去の軌跡は $\mathbf{X}_i = \{(x_i^t, y_i^t) | t = 1, \dots, t_{obs}\}$ 、 i 番目の歩行者と周囲の障害物との距離情報は $\mathbf{S}_i = \{s_i^t | t = 1, \dots, t_{obs}\}$ 、予測する i 番目の歩行者の将来位置は $\mathbf{Y}_i = \{(x_i^t, y_i^t) | t = t_{obs}+1, \dots, t_{obs}+t_{pred}\}$ と表される。ただし、 t_{obs} 、 t_{pred} はそれぞれ観測時間と予測時間を表す。

本手法で用いられるネットワークは、主にエンコーダモジュールとデコーダモジュールの2つのモジュールで構成される。全体的なネットワークの構造を図1に示す。エンコーダでは、入力となる各時刻における歩行者の過去の位置は、全結合層 ϕ_e により状態ベクトルに埋め込まれる。そして、状態ベクトルは時系列データとして LSTM (Hochreiter and Schmidhuder, 1997) に入力され、歩行者が過去に辿ってきた軌道を潜在的に表現する隠れ状態ベク

トルを得る。エンコーダでの処理を式 (3) と式 (4) に示す。

$$e_{ei}^t = \phi_e(x_i^t, y_i^t; W_{ee}) \quad (3)$$

$$h_{ei}^t = \text{LSTM}(h_{ei}^{t-1}, e_{ei}^t; W_{le}) \quad (4)$$

W_{ee} , W_{le} はそれぞれ重みを表し、重みは全ての歩行者ノードで共有される。

時刻 $t = 1$ から $t = t_{obs}$ における i 番目の歩行者と障害物の距離の時系列情報 S_i はエンコーダと同様、全結合層により埋め込まれ、LSTM により歩行者と周囲の障害物との距離情報の時間変化を潜在的に表現する特徴量を取得する。これにより、歩行者が周囲の物体との距離の変化を知覚し、行動に反映させていることの再現を図る。距離情報の処理を式 (5) と式 (6) に示す。

$$e_{si}^t = \phi_s(S_i^t; W_{es}) \quad (5)$$

$$h_{si}^t = \text{LSTM}(h_{si}^{t-1}, e_{si}^t; W_{ls}) \quad (6)$$

時刻 $t = t_{obs}$ におけるデコーダへの入力は、観測された歩行者位置をエンコーダとは異なる重みを持つ全結合層 ϕ_d で埋め込んだ状態ベクトルである。そして、エンコーダの出力 h_{ei} , h_{si} , GAT の出力 g_i を結合し、多層パーセプトロン (MLP) により変換した特徴量によりデコーダの LSTM の隠れ状態ベクトルを初期化する。デコーダの処理を式 (7), 式 (8), 式 (9), に示す。

$$e_{di}^t = \phi_d(x_i^t, y_i^t; W_{ed}) \quad (7)$$

$$g_i^t = \text{GAT}(h_{ei}^t) \quad (8)$$

$$h_{di}^t = \text{LSTM}(\text{MLP}(h_{di}^{t-1}, h_{si}^{t-1}, g_i^t), e_{di}^t; W_{le}) \quad (9)$$

デコーダの出力は全結合層 ϕ_{mnd} を通過し、式 (10) により各歩行者の予測時刻における位置を表す確率分布に変換される。

$$\hat{Y}_i^t = \phi_{mnd}(h_{di}^t; W_{mnd}) \quad (10)$$

ただし、 W_{mnd} は全結合層の重み、 $\hat{Y}_i^t$ は確率分布であり、式 (11) で表される。

$$\hat{Y}_i^t = [\mu_x, \mu_y, \sigma_x, \sigma_y, \rho]^t \quad (11)$$

ここで、 μ は平均、 σ は標準偏差、 ρ は相関係数である。

2.3 Graph Attention Network

衝突を回避するために他の歩行者の重要度を割り当て、相互作用の相対的な重みを算出するために Graph Attention Network (Veličković et al., 2017) を用いる。Graph Attention Layer の入力、隠れ状態ベクトル ($h = h_1, h_2, \dots, h_N, h_i \in R^F$) である。ただし、 N はノード数、 F はノードの次元数を表す。ノード i に対するノード j の係数は式 (12) と式 (13) によって求められる。

$$e_{i,j}^t = a(W h_i^t, W h_j^t) \quad (12)$$

$$\alpha_{i,j} = \text{softmax}(e_{i,j}) \quad (13)$$

Table 1 Details of the simulation environments

	Square	Corner	Obstacle	Crossroads
Number of static obstacles	0	0	4	0
Number of pedestrians	10	3	7	8
Movable range of pedestrians (x / y) [m]	20 / 20	8 / 25	5 / 20	15 / 24

ただし, $W \in \mathbf{R}^{F' \times F}$ は線形変換の重み行列, $a: \mathbf{R}^{F'} \times \mathbf{R}^{F'} \rightarrow \mathbf{R}$ は shared attention mechanism である. ノード i に対する GAT の最終的な出力は式 (14) で示される.

$$g_i^t = \sigma(\sum_{j \in N} \alpha_{i,j} W h_j) \quad (14)$$

ただし, σ は非線形関数であり, 本手法では ELU (Clevert et al., 2016) を用いた.

2.4 学習方法

提案するネットワークの学習手法の詳細について述べる. デコーダにより予測された歩行者の確率分布は正規分布に従うと仮定し, 負の対数尤度関数 E_i を損失関数として適用する. 損失関数を式 (15) に示す.

$$E_i = \sum_{t=t_{obs}+1}^{t_{obs}+t_{pred}} \log \left(P(\mathbf{x}_i^{(t)} | \boldsymbol{\mu}_i^{(t)}, \boldsymbol{\Sigma}_i^{(t)}) \right) \quad (15)$$

ただし, $\boldsymbol{\mu}$ は分布の平均, $\boldsymbol{\Sigma}$ は分散共分散行列である. この損失関数を最小化するようにネットワークの重みを更新することで, 学習を進める.

学習に用いたパラメータは, LSTM の隠れ状態ベクトルの次元を 30, 埋め込みに用いた全結合層の次元を 16 とした. また, 多層パーセプトロンの中間層の次元を 64 とし, 活性化関数には ReLU を用いた. 最適化には Adam (Kingma and Ba, 2015) を用い, バッチサイズは 32, 学習率は 0.0001 とし, 200 エポック学習を行った. また, ビームモデルにおけるカーネルサイズは 9, 最大距離は 5.0[m], 歩行者の視野角を 180°, 分解能を 1° とした.

3. 実 験

3.1 実験概要

実験にはシミュレータ上で作成されたデータセットと, 公開されている歩行者の軌道を含むデータセットを用いる. これらのデータセットでネットワークの学習を行い, テストデータに対する予測精度の評価を行う. ネットワークの入力データには, 1 ステップを 0.5 秒とした観測時間 t_{obs} ステップ分の歩行者位置と, 仮想ビームモデル (Thrun et al., 2005) により算出した歩行者と障害物の距離情報を用い, 正解ラベルとしてそれに対応した予測時間 t_{pred} ステップ分の歩行者の将来位置を用意し, これを 1 セット分のデータとする. 本実験では, $t_{obs} = 8$ (4 秒), $t_{pred} = 8$ (4 秒) とした. 予測結果の評価には Final Displacement Error(FDE) と Average Displacement Error(ADE) を指標として用いた. FDE は予測結果と真値の最終時刻 ($t = t_{obs} + t_{pred}$) における距離を表し, ADE は時刻 $t = t_{obs}$ から $t_{obs} + t_{pred}$ までの各時刻における予測結果と真値の距離の平均を表す. FDE を式 (16), ADE を式 (17) に示す.

$$\text{FDE} = \sqrt{(x_{t_{pred}} - \hat{x}_{t_{pred}})^2 + (y_{t_{pred}} - \hat{y}_{t_{pred}})^2} \quad (16)$$

$$\text{ADE} = \frac{\sum_{t=t_{obs}}^{t_{pred}} \sqrt{(x_t - \hat{x}_t)^2 + (y_t - \hat{y}_t)^2}}{t_{pred}} \quad (17)$$

ただし, $(\hat{x}_t, \hat{y}_t), (x_t, y_t)$ は時刻 t における歩行者位置の真値と予測結果である. また, 提案手法を評価するために,

Fig. 2 The simulation environments are shown. There are four simulation environments, the upper left figure is an environment assuming a square, the upper right figure is an environment assuming a corner, the lower left figure is an environment with multiple static obstacles, and the lower right figure is an environment assuming a crossroads.

以下の6つの手法を用いて予測を行い、精度の比較を行う。なお、Social STGCNN (Mohamed et al., 2020) は評価を行う際に予測分布から複数の軌道をサンプルし、最も誤差の小さい軌道を評価に使用することが提案されているが、本実験では、他の手法との公平性を保つため、20本サンプルした場合に加え、最も尤度が大きい軌道1本のみを用いる場合の2つで評価を行う。また、提案手法のネットワークにおいて障害物情報と GAT (Veličković et al., 2017) を用いる有用性を示すため、障害物情報を用い、GAT を用いない場合と、GAT を用い、障害物情報を用いない場合の予測結果についての比較も行う。

- 1) 歩行者が速度一定で直進
- 2) Social LSTM (Alahi et al., 2016)
- 3) Social STGCNN (Mohamed et al., 2020)
- 4) GAT
- 5) Beam Model
- 6) GAT + Beam Model (提案手法)

3.1.1 シミュレーション環境での実験

本実験では、ロボット用3次元動力学シミュレータである Gazebo を用いて作成した、壁などの静的障害物が存在するシミュレーション環境を用いる。このシミュレーション環境で各歩行者と障害物との距離情報を含むデータセットを作成し、これを用いて学習したモデルを用いて精度評価を行う。シミュレーションには、

- A) 障害物の存在しない広場を想定した環境 (図 2(a))
- B) 曲がり角を想定した環境 (図 2(b))
- C) 静的障害物が多数存在する環境 (図 2(c))
- D) 十字路を想定した環境 (図 2(d))

の4つの環境を用いた。各シミュレーション環境の静的障害物の数、歩行者の数、歩行者の歩行可能範囲を表1に示す。各環境における歩行者の数と障害物は、歩行者が周囲との相互作用を伴うことと、スタックすることなく自然に目的地にたどり着くことが両立できる条件を設定している。シミュレータ内の各環境において Social Force Model (Helbing and Molnar, 1995) に基づく歩行者モデルをランダムに生成した目的地に向けて速度 $0.8 - 1.2\text{m/s}$ で移動させ、環境内に配置した差動二輪ロボットの3次元モデルを用いてデータの取得を行った。実験データには

Table 2 Quantitative result of ADE and FDE in the simulation environment

Methods \ Datasets	Square		Obstacle		Corner		Crossroads	
	ADE[m]	FDE[m]	ADE[m]	FDE[m]	ADE[m]	FDE[m]	ADE[m]	FDE[m]
Linear	0.511	0.812	0.563	0.899	1.12	1.96	0.452	0.726
Social LSTM	0.406	0.768	0.523	0.831	0.440	0.824	0.339	0.667
Social STGCNN(1)	0.574	1.04	0.556	0.976	0.623	1.17	0.482	0.892
Social STGCNN(20)	0.378	0.592	0.501	0.488	0.405	0.672	0.299	0.422
GAT	0.351	0.675	0.512	0.829	0.538	1.09	0.367	0.712
beam model	0.333	0.641	0.456	0.810	0.318	0.602	0.332	0.594
GAT + Beam Model(ours)	0.305	0.548	0.455	0.804	0.347	0.637	0.312	0.576

Table 3 Quantitative result of ADE and FDE in the real world

Methods \ Datasets	THÖR	
	ADE[m]	FDE[m]
Linear	0.089	0.143
Social LSTM	0.0693	0.132
Social STGCNN(1)	0.0805	0.141
Social STGCNN(20)	0.0708	0.0641
GAT	0.0580	0.121
Beam Model	0.0563	0.115
GAT + Beam Model(ours)	0.0555	0.103

各シミュレーション環境において作成したデータのうち, [A]4200 セット [B]3500 セット [C]3000 セット [D]3000 セットを用い, テストデータには [A]700 セット [B]500 セット [C]500 セット [D]600 セットを用いた.

3.1.2 実環境における実験

本実験では, 公開されている実環境で作成されたデータセットの, THÖR (Rudenko et al., 2020) を用いて評価を行う. THÖR は人間の動き, 視線データ, 3D LiDAR によって収集されたセンサデータを含むデータセットであり, いくつかの障害物が存在する室内環境において歩行者同士が相互作用を伴う動作を行うシナリオが設定されている. 歩行者の速度は $0.32 - 1.3\text{m/s}$ で, 歩行者の位置はモーションキャプチャにより 100Hz で記録されており, 本実験ではこのデータから 0.5 秒間隔でデータを抽出し, 軌跡を生成した. データセットは取得したデータ 20000 セットのうち, 14000 セットを学習データとし, 6000 セットをテストデータとした.

3.2 実験結果

シミュレーション環境におけるテストデータの予測に対する誤差を表 2 に示す. 表において, Social STGCNN に付いている番号は, 分布からサンプルした軌道の数を表している. また, GAT は提案手法において障害物情報を用いなかった場合の予測結果, Beam Model は GAT を用いなかった場合の予測結果を表している. 提案手法の予測は, 広場を想定した環境, 障害物の存在する環境において最も ADE が小さく, 広場を想定した環境において最も FDE が小さい結果となった.

障害物の存在する環境において, 提案手法の比較手法の中で最も ADE が小さいのは Social STGCNN であり, 0.501 であった. これに対し, 提案手法の ADE は 0.455 であり, 比較手法に対し 9.2% 小さい結果となった. また, 曲がり角を想定した環境においても, Social STGCNN の ADE 0.405 に対し, 提案手法の ADE は 0.347 であり, 14.3% 小さい結果となった. 曲がり角を想定した環境においては, GAT を用いず Beam Model のみを用いた場合の方が ADE, FDE ともに小さい結果となったが, このことから, 曲がり角のように障害物により大きく進路を変更

Fig. 3 An example of the prediction result of the pedestrian trajectory is shown. In the figure, the green dots indicate the observed past trajectories, the blue dots indicate the true values of the future positions, the red dots indicate the prediction results of the future trajectories, and the black areas indicate static obstacles.

するような状況では、障害物との距離の変化が大きく影響しているということが考えられる。また、広場を想定した環境において、GATのみを用いた場合の ADE は 0.351, Beam Model のみを用いた場合の ADE は 0.333, 両方を用いた場合の ADE は 0.305 という結果となった。歩行者が他の歩行者などの動的障害物を回避するような場面においても、各歩行者間の関係性を GAT により学習するとともに、障害物との距離の変化も加味することでより正確な予測を行うことが示唆されている。

Social LSTM の ADE に対する FDE の値は、障害物の存在する環境において 1.59 倍、十字路を想定した環境において 1.97 倍であり、提案手法はそれぞれ 1.85 倍, 1.77 倍となった。これに対し, Social STGCNN の場合は 1.41 倍, 0.97 倍であり, 他の手法に比べて ADE に対して FDE が小さい傾向が見受けられた。この原因として, 2 つの

ことが考えられる。1つ目は Social STGCNN の評価は、生成した複数の軌道から最も誤差の小さいものを用いているため、予測結果の終端位置が真値に近くなったということが挙げられる。2つ目は本手法が従来の手法で用いられていた RNN ではなく CNN を用いているということである。RNN を用いて予測を行う場合、次の時刻の歩行者位置を算出する際に、一時刻前の予測結果を用いるため、誤差が積算しやすい傾向があるが、CNN を用いて予測を行う場合、シーケンス全体を予測するため、RNN のように誤差が積算しづらいということが考えられる。

実環境データに対する誤差を表 3 に示す。ADE は提案手法が一番小さく、FDE は Social STGCNN の誤差が最も小さい結果となった。Social LSTM は、速度一定に比べて、ADE 22.1%、FDE 7.7%減少しているが、提案手法は、ADE 37.6%、FDE 28.0%と大きく減少しており、提案手法が実環境においても有効であることが分かる。

実環境データセットとシミュレーション環境の誤差を比較すると、全体的に実環境データの方が小さい値となったが、これは、THÖR データセットの歩行パターンがシミュレーション環境に比べ、直進するものが多かったことや、その場で停滞して、動きが少ない歩行者がいたことなどが原因として考えられる。

図 3 に予測結果の例を示す。図中の緑で示された部分が、歩行者の過去の軌跡、青で示された部分が歩行者の将来位置の真値、赤で示された部分が予測された将来の軌跡、黒で示された部分が静的障害物を表している。また、Social STGCNN 予測結果において、予測軌道の周囲の黒く表示されている部分は確率分布を表しており、最も尤度が高い軌道が表示されている。図 3(a)、図 3(d)、図 3(g)において、速度一定、Social LSTM、Social STGCNN の予測結果では進路上の障害物上を通過するような軌道を予測してしまっている。一方、図 3(j)ではどちらの歩行者も進路上の障害物を回避する動作を予測することができている。また、図 3(b)、図 3(e)、図 3(h)において、歩行者は曲がり角で進路を右に変更するような動作をとっているが、これを予測することができていないのに対し、提案手法による予測結果(図 3(k))では、歩行者が曲がる動作を予測することができている。また、図 3(c)、図 3(f)、図 3(i)、図 3(l)において、速度一定での予測では、上から下に向かって歩いている歩行者が、反対側から進行してきた歩行者を回避する動作を予測することができていないが、Social LSTM による予測と、提案手法による予測では予測することができている。これらのことから、提案手法による予測では、他者や、周囲の障害物からの影響を考慮した予測をすることができていると言える。

4. 結 言

本論文では、歩行者の過去の動作と、周囲の環境に存在する障害物との距離情報を用いて歩行者の将来の軌道を予測する手法を提案した。シミュレーション環境において取得したデータを用いた予測では、従来手法を上回る精度を示した。また、公開されている実環境で取得されたデータセットを用いた予測でも、従来手法を上回る精度を示し、従来手法では困難であった障害物を回避する動作を予測することができていることが確認できた。しかし現状の手法では、歩行者や障害物が多数存在する環境において、オクルージョンにより占有格子地図上の欠損部が多くなり、歩行者と障害物の距離情報を正しく取得することができないといった問題がある。今後の発展としては、画像から得られる歩行者の視線や姿勢などの情報を考慮した上で予測を行うことで、より様々な状況に適応した予測を行うことができると考えられる。

謝 辞

本研究の一部は、NEDO 次世代人工知能・ロボット中核技術開発の支援を受けた。ここに篤く御礼申し上げる。また、本研究は明治大学自律型ロボット研究クラスターの下で実施された。

文 献

- Alahi, A., Goel, K., Ramanathan, V., Robicquet, A., Fei-Fei, L. and Savarese, S., Social lstm: human trajectory prediction in crowded spaces, Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)(2016), pp.961-971.
- Clevert, D., Unterthiner, T. and Vandergheynst, P., Fast and accurate deep network learning by exponential linear units(elus), Proceedings of the International Conference on Learning Representations (ICLR)(2016).

- Gupta, A., Johnson, J., Fei-Fei, L., Savarese, S. and Alahi, A., Social gan: socially acceptable trajectories with generative adversarial networks, *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*(2018).
- Helbing, D. and Molnar, P., Social force model for pedestrian dynamics, *Physical review E*, Vol. 51, No. 5 (1995), pp.4282-4286.
- Hochreiter, S. and Schmidhuber, J., Long short-term memory, *Neural computation*, Vol.9, No.8 (1997), pp.1735-1780.
- Huang, Y., Bi, H., Li, Z., Mao, T. and Wang, Z., Stgat: modeling spatialtemporal interactions for human trajectory prediction, *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*(2019), pp.6272–6281.
- Kingma, D. P. and Ba, J. L., Adam: a method for stochastic optimization, *Proceedings of the International Conference on Learning Representations (ICLR)* (2015), pp.1-15.
- Mohamed, A., Qian, K., Elhoseiny, M. and Claudel, C., Social-stgcnn: a social spatio-temporal graph convolutional neural network for human trajectory prediction, *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*(2020).
- Rudenko, A., Kucner, T. P., Swaminathan, C. S., Chadalavada, R. T., Arras, K. O., and Lilienthal, A. J., Thör: human-robot navigation data collection and accurate motion trajectories dataset, *IEEE Robotics and Automation Letters*, Vol.5, No.2 (2020).
- Sadeghian, A., Kosaraju, V., Sadeghian, A., Hirose, N., Rezatofighi, H. and Savarese, S., Sophie: an attentive gan for predicting paths compliant to social and physical constraints, *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*(2019).
- Styles, O., Guha, T. and Sanchez, V., Multiple object forecasting: predicting future object locations in diverse environments, *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision (WACV)*(2020), pp.690-699.
- 高梨宏之, 阿部勝哉, 道辻洋平, 小竹元基, ポンサトーン・ラクシンチャランサク, 林隆三, 路肩の障害物回避における歩行経路の確率的予測モデル構築, *日本機械学会論文集* Vol.71, No.855 (2017), DOI:10.1299/transjsme.17-00224.
- Thrun, S., Burgard, W. and Fox, D., *Probabilistic Robotics*, The MIT press (2005).
- Veličković, P., Cucurull, G., Casanova, A., Romero, A., Liò, P. and Bengio, Y., Graph attention networks, *Proceedings of the International Conference on Learning Representations (ICLR)*(2017).
- Yagi, T., Mangalam, K., Yonetani, R. and Sato, Y., Future person localization in first-person videos, *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*(2018).

References

- Alahi, A., Goel, K., Ramanathan, V., Robicquet, A., Fei-Fei, L. and Savarese, S., Social lstm: human trajectory prediction in crowded spaces, *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*(2016), pp.961-971.
- Clevert, D., Unterthiner, T. and Vandergheynst, P., Fast and accurate deep network learning by exponential linear units(elus), *Proceedings of the International Conference on Learning Representations (ICLR)*(2016).
- Gupta, A., Johnson, J., Fei-Fei, L., Savarese, S. and Alahi, A., Social gan: socially acceptable trajectories with generative adversarial networks, *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*(2018).
- Helbing, D. and Molnar, P., Social force model for pedestrian dynamics, *Physical review E*, Vol. 51, No. 5 (1995), pp.4282-4286.
- Hochreiter, S. and Schmidhuber, J., Long short-term memory, *Neural computation*, Vol.9, No.8 (1997), pp.1735-1780.
- Huang, Y., Bi, H., Li, Z., Mao, T. and Wang, Z., Stgat: modeling spatialtemporal interactions for human trajectory prediction, *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*(2019), pp.6272–6281.
- Kingma, D. P. and Ba, J. L., Adam: a method for stochastic optimization, *Proceedings of the International Conference on Learning Representations (ICLR)* (2015), pp.1-15.

- Mohamed, A., Qian, K., Elhoseiny, M. and Claudel, C., Social-stgcnn: a social spatio-temporal graph convolutional neural network for human trajectory prediction, Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)(2020).
- Rudenko, A., Kucner, T. P., Swaminathan, C. S., Chadalavada, R. T., Arras, K. O., and Lilienthal, A. J., Thör: human-robot navigation data collection and accurate motion trajectories dataset, IEEE Robotics and Automation Letters, Vol.5, No.2 (2020).
- Sadeghian, A., Kosaraju, V., Sadeghian, A., Hirose, N., Rezaatofghi, H. and Savarese, S., Sophie: an attentive gan for predicting paths compliant to social and physical constraints, Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)(2019).
- Styles, O., Guha, T. and Sanchez, V., Multiple object forecasting: predicting future object locations in diverse environments, Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision (WACV)(2020), pp.690-699.
- Takanashi, H., Abe, K., Michitsuji, Y., Shino, M., Raksincharoensak, P. and Hayashi, R., Stochastic prediction model for obstacle avoidance route of pedestrian, Transactions of the JSME (in Japanese), Vol.83, No.855 (2017), DOI:10.1299/transjsme.17-00224.
- Thrun, S., Burgard, W. and Fox, D., Probabilistic Robotics, The MIT press (2005).
- Veličković, P., Cucurull, G., Casanova, A., Romero, A., Liò, P. and Bengio, Y., Graph attention networks, Proceedings of the International Conference on Learning Representations (ICLR)(2017).
- Yagi, T., Mangalam, K., Yonetani, R. and Sato, Y., Future person localization in first-person videos, Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)(2018).