

人物自動追従ロボットのための 機械学習による歩行者経路予測手法

赤羽根 里奈^{1,a)} 加藤 由花¹

概要：本稿では、人物追従ロボットの追従精度向上のために、機械学習によりターゲットとなる歩行者の未来の移動経路を予測する手法を提案する。提案手法は、公開されているデータセットを用いて事前に予測器を構築しておくことにより、予測時にはロボット単体での経路予測を実現する点に特徴がある。ここでは、測域センサー単体で複数歩行者の移動軌跡を観測した上で、経路予測には、既存の系列データ予測手法である LSTM (Long Short-Term Memory) ベースの手法を用いる。特に、人と人との相互作用をモデルに組み込んだ手法を用いることで、混雑した環境での経路予測を目指す。さらに、学習済み予測器による予測の有効性を検証するために、2種類のデータセットから構築した予測器を用いた実験を行い、データセットの違いが予測結果に与える影響を評価する。

キーワード：人追従ロボット、歩行者経路予測、学習済み予測器、データセット、LRS

1. はじめに

サービスロボットの普及に伴い、家庭や職場、公共施設などで人と共存するロボットに対する期待が高まっている。これらのロボットの一種に、人物自動追従ロボットがある。これは、ターゲットとなる歩行者を識別し、追従し続けるロボットであり、重い荷物の搬送、屋内外の施設案内、医療現場や工場等での作業補助等、様々な場面での利用が期待されている。我々の研究グループでも、これまで、深度カメラにより人の骨格形状を取得することで人の初期位置を決定し、その後、測域センサー (Laser Range Scanner: LRS) を用いてターゲットを追尾するロボットの研究を進めてきた [1]。ここでは、主に公共施設 (博物館等) での利用を前提に、センシング結果とロボットの移動モデルに基づくベイズフィルタにより人の追従を行う手法を実現したが、障害物による遮蔽等でロボットが人を見失うケースが多く、特に混雑環境での追従が困難であるという課題が残った。

本稿では、この課題を解決するために、機械学習によりターゲットとなる歩行者の未来の経路を予測し、その結果を追従制御に利用することを考える。人間は、ターゲットとなる人を追従する場合、たとえターゲットとの間に障害物が割り込んだとしても、未来の人の位置を予測すること

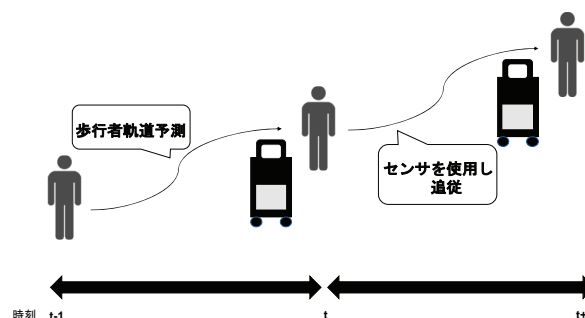


図 1 提案手法の利用イメージ。

で再びターゲットを検出し、追従することが可能である。提案手法は、この仕組みを追従ロボットに組み込むことを目指すものである。手法の利用イメージを図 1 に示す。

歩行者経路予測については、これまで多くの研究開発が行われてきた。一つは、歩行者の移動傾向を数理モデルとして定式化し、このモデルに従い未来の経路を予測するものである。代表的なモデルに Helbing の Social Force Model (SFM) [2] がある。もう一つは、大量の観測データを利用し、対象環境における歩行者の移動傾向を機械学習により推定するものである。前者は、適切なモデル化が行われれば適用は容易だが、複雑な環境を対象とした場合には考慮すべきパラメータ数が膨大になり、モデル構築自体が困難になってしまう。一方、後者は、適切な訓練データを用意できれば、比較的複雑な環境のモデル化 (予測器の構築) が可能である。本稿では、ロボットが人を見失う

¹ 東京女子大学 大学院理学研究科
Suginami, Tokyo 167-8585, Japan

^{a)} d19m201@cis.twcu.ac.jp

ような複雑な環境を対象とするため、後者の手法を採用することにした。近年、移動データを系列データとして学習させる研究が盛んに行われており、様々な予測手法が提案されている [3] [4]。本稿では、これらの中から、環境ごとに適した手法を選んで利用する。

ここで、予測器の構築には訓練用に大量の移動軌跡データが必要になるが、一般の公共空間では、サービス提供前に十分な量のデータを収集しておくことが難しい。また、環境側へのセンサの設置も困難である場合が多い。我々はこれまで、公開されているデータセットを用いて事前に構築しておいた予測器を用いて、歩行者の到達地点を予測する手法を提案してきた [5]。本稿でも同様の考え方を適用し、予測対象とは異なる環境で、異なるセンサーを用いて構築した予測器を用いて、移動経路の予測を行う。

本稿の貢献は以下の2点である。

- 追従ロボットのための経路予測手法として、環境ごとに適切なデータセットと予測モデルを選択する手法を提案する。
- 予測器を構築するデータセットの違いが、予測結果に与える影響を評価する。

2. 関連研究

本稿では、歩行者の過去の移動経路から未来の移動経路を予測し、それをロボットの人物追従に利用する。歩行者経路予測は、近年、特にビジョン研究の分野で活発に取り組まれている研究課題であり [6]、前述した研究 [3][4] の他、Trajectory Forecasting Challenge^{*1}なども行われている。本章では、それら経路予測に関する研究、および代表的なロボット人追従手法に関する研究について説明する。

2.1 歩行者経路予測に関する研究

歩行者経路予測については、まず、古くから研究が行われてきたベイズフィルタに基づく一連の研究が存在する。これらの手法では、ターゲットのトラッキングを目的に、短期的な経路予測が行われる。例えば、観測された歩行速度を利用し、近接時の人と人の相互作用を移動モデルに組み入れることで拡張カルマンフィルタを構成する手法 [7]、複数の手がかり (Appearance, Motion, Interaction) に基づきそれぞれ独立に構築された予測器 (LSTM (Long Short-Term Memory) [8] により構築) を組み合わせ、次周期の複数ターゲットの位置を予測する手法 [9]、マルコフ決定過程を用いてターゲットの状態 (Active, Inactive, Tracked, Lost) を推定し、状態ごとに処理内容を定義する手法 [10]、CNN (Convolutional Neural Network) に基づく歩行者検出と組み合わせることで、リアルタイムで追跡候補を予測する手法 [11] などがある。これらの手法は、

ターゲット再検出のための様々な工夫がなされているが、いずれも、ある程度以上の長期間の予測は困難である。

そのため、近年では、RNN (Recurrent Neural Network) ベースの手法を用いて予測器を構築し、系列データとして経路予測を行う手法も提案されている。各自の移動経路 (系列データ) を LSTM でモデル化し、隠れ層の状態を共有することで人と人との相互作用を表現する Social LSTM [3] や、これを GAN (Generative Adversarial Networks) を用いた手法に拡張し、複数人の経路を同時に予測 (生成) 可能にした Social GAN [4] などがある。

本稿では、これらの系列データ予測手法を利用し、未来の移動経路を予測する。

2.2 ロボット人追従手法に関する研究

ロボットによる人追従手法としては、LRS によるセンシング結果を利用する手法がよく用いられる。精度が高く、日照など環境の影響を受けにくいこと、単体でターゲットの大きさ、位置、移動方向の推定が可能であることから、様々な環境で利用されている。カメラを用いる手法と異なり、視覚情報の取得 (一般物体認識等) はできないが、様々な工夫により、ターゲットを検出した上で追従を実現している。例えば、検出対象 (人の足の見え方) のテンプレートを用意しておき、テンプレートマッチングにより位置を推定する手法などが提案されている [12]。追従にはベイズフィルタ (Particle Filter) が用いられている。

本稿でも、ロボットの人物追従にはこれらの手法を用いるが、前述の系列データ予測手法と組み合わせることにより、追従の精度向上を目指す。

3. 提案手法

3.1 手法の概要

本稿では、博物館等の公共空間に設置され、人に追従しながら案内を行うロボットを対象に、追従精度を向上させるための歩行者経路予測手法を提案する。ここでは、公開されているデータセットを用いて事前に予測器を構築しておき、追従ロボットに搭載された LRS によるセンシング結果を用いて、歩行者の未来の経路を予測する。提案手法の概要を図 2 に示す。手法は、予測器を構築する学習フェーズ (オフライン) と、経路を予測する予測フェーズ (オンライン) の2つのフェーズに分かれており、以下の手順に従って予測を行う。

- 学習フェーズ
 - (i) データセットから訓練データを生成する。
 - (ii) 訓練データを用いて予測器を構築する。
- 予測フェーズ
 - (i) 移動経路データを計測する。
 - (ii) 計測データから予測器への入力データを生成する。
 - (iii) 移動経路を予測する。

^{*1} <http://trajnet.stanford.edu>

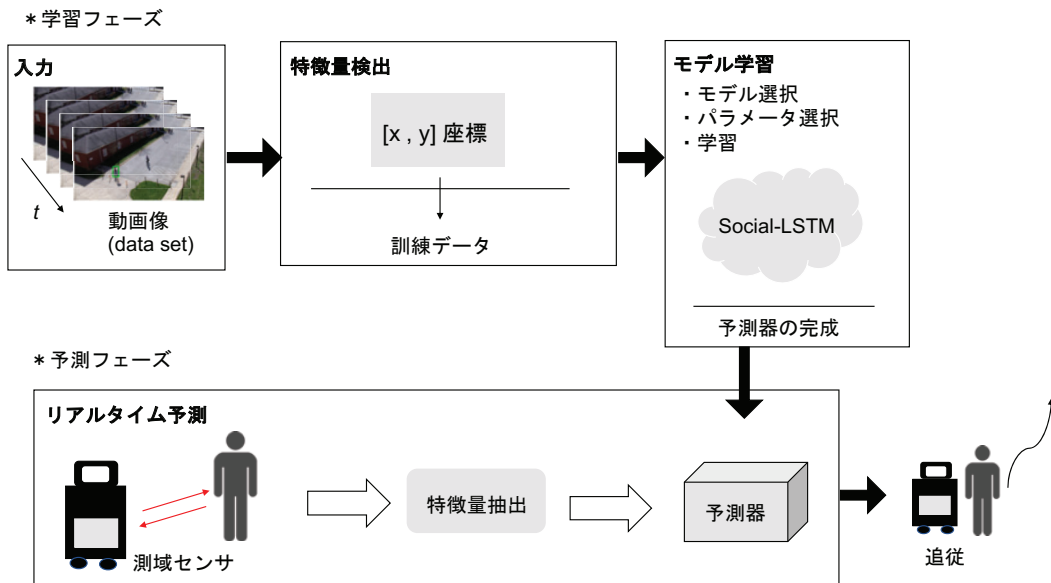


図 2 提案手法の概要.

3.2 学習フェーズ

3.2.1 訓練データの生成

本稿では、環境の真上から撮影された鳥瞰視点の映像データセットであり、経路予測に用いられることの多い UCY Dataset [13] と ETH Dataset [7], および複数センサーのデータを含む PETS 2009 Benchmark Data*2を対象データセットとして用いる。これらは市街地の歩行者を撮影したシーンからなるデータセットであり、歩行者のみが存在するシーン（車や自転車等は存在しない）を撮影した動画画像で構成される。本稿では、博物館等を利用シーンとして想定するため、歩行者のみが存在するデータセットを利用することにした。なお、動画画像そのものではなく、歩行者の座標データが数値として抽出されているデータセットを利用している。シーン数は UCY が 3, ETH が 2, PETS は 3, 歩行者数は UCY が 786 人, ETH が 750 人, PETS は不明（最大密度は 40 人/m²）である。

データセットは、以下の情報を 1 レコードとして、映像フレームごとに観測された歩行者 ID 分のデータから成る。

- フレーム ID（タイムスタンプに相当）
- 歩行者 ID
- y 座標, x 座標 [m]

つまり、あるレコードを取り出すと、ある歩行者のある映像フレーム（時刻）における位置座標を得ることができる。なお、各歩行者の移動経路は、20 フレーム分ずつ抽出されている。データセットの一例を表 1 に示す。

訓練データは、データセットに含まれるデータをそのまま利用する。ここでは、対象となる環境と類似したシーンを選択し、学習用の訓練データとする。

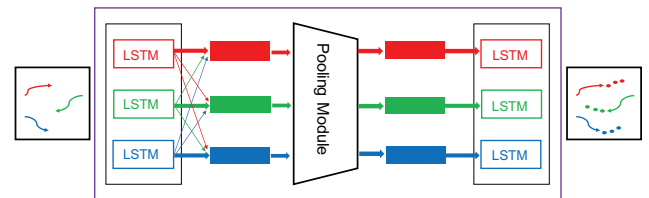


図 3 Social-LSTM のイメージ. 文献 [4] を参考に作成.

3.2.2 予測器の構築

予測器の構築には、系列データとして経路を予測する LSTM ベースの手法を用いる。本稿では、2.1 節で取り上げた Social LSTM を用いることにした。Social LSTM では、ある領域内の全ての歩行者が隠れ層の状態を共有することにより、混雑状態での人と人との相互作用（お互いに衝突を避けようと移動する行動等）を予測モデルに組み込んでいる。手法のイメージを図 3 に示す。

文献 [3] では、歩行者間で状態を共有しない手法（Vanilla LSTM）および隣接する歩行者同士のみで状態を共有する手法（Obstacle LSTM）との比較が行われており、データセットを用いた実験により、Social LSTM の有効性が示されている。特に、本稿で対象とする比較的混雑した環境での性能差が大きいことから、Social LSTM を選択すること

表 1 データセットの一例

フレーム ID	ユーザ ID	y 座標	x 座標
0	5	1.61	0.94
10	5	1.62	0.95
⋮	⋮	⋮	⋮
180	5	1.55	0.98
190	5	1.52	0.83
0	6	1.02	1.13
⋮	⋮	⋮	⋮

*2 <http://www.cvg.reading.ac.uk/PETS2009/data.html>

にした。なお、本稿では、予備検証として ETH に含まれる学習用データセットのデータを用いて予測器を構築し、ETH および UCY に含まれる全シーンのデータを用いて予測を行った。その結果を表 2 に示す（各シーンごとに予測を行い、全シーンに対する結果の平均値を求めた）。評価尺度は、予測経路と真の経路との各フレーム座標の平均誤差（Average Error）、および最終的な到達地点の予測値と真値の差（Final Error）である。

この結果より、Social LSTM の優位が確認されたが、同時に、隣接する歩行者のみで状態を共有する Obstacle LSTM の効果はほとんど見られないこともわかった。

3.3 予測フェーズ

3.3.1 データの計測

予測フェーズでは、LRS の計測範囲内に入った人の位置追跡を行い、10Hz 周期で LRS から人までの距離と角度を計測する。人位置追跡には、文献 [14] に記載されたものと同様の方法を用いる（ただし、文献 [14] では人の足首を追跡対象としているが、本稿では胴体を追跡対象とする）。ここでは、追跡対象の位置を確率分布として扱い、追跡対象に応じた遷移モデル（人の移動）と、センサデータに基づく観測モデル（LRS により測定される距離）から、再帰的に追跡対象の状態量に関する確率密度関数を推定する。人位置（追跡器の位置）の確率分布には、Particle Filter を用いる。Particle Filter では、確率分布は標本（Particle）の集合として定義され、Particle を重みに従ってリサンプリングすることで、事後確率（同じく Particle の集合として定義される）が生成される。

歩行者の位置は、LRS により計測される距離と角度を、極座標形式から直交座標形式に変換することで求める。これらの値は、いずれも、ロボットに固定されたローカル座標系のものである。この結果から、単位時間ごとに、計測範囲内に存在する歩行者ごとの位置データ $(\hat{x}_t^{(i)}, \hat{y}_t^{(i)})$ が算出される。ここで t は時刻、 i は歩行者を識別する ID である。

3.3.2 入力データの生成

前項の手順で得られた計測結果から、予測器への入力データを生成する。予測器には、対象環境に固定されたワールド座標系での歩行者の位置座標の点列を入力する必要があるため、データ形式を合わせる処理を行う。具体的には、前項で取得される $(\hat{x}_t^{(i)}, \hat{y}_t^{(i)})$ を、20 フレーム $(0 \leq t \leq 19)$ ごとに観測された歩行者の人数分（ i の最大値分）収集し、ローカル座標からワールド座標に変換した上で、表 1 に示

すように、1 歩行者、1 フレームを 1 レコードとしてデータ列を生成する。なお、ワールド座標系での時刻 t における追従ロボットの自己位置の座標 $(X_t^{(R)}, Y_t^{(R)})$ および姿勢 φ_t は、他の手法（環境地図を用いた自己位置推定手法など）により別途算出されているものと仮定する。ここで $(X_t^{(R)}, Y_t^{(R)})$ はロボットの車軸の中心の座標、 φ_t はワールド座標とロボットの正面が成す角度である。

すると、ワールド座標系における、時刻 t における歩行者 i の位置座標 $(x_t^{(i)}, y_t^{(i)})$ は、

$$\begin{pmatrix} x_t^{(i)} \\ y_t^{(i)} \end{pmatrix} = \begin{pmatrix} \cos \varphi_t & -\sin \varphi_t \\ \sin \varphi_t & \cos \varphi_t \end{pmatrix} \begin{pmatrix} \hat{x}_t^{(i)} \\ \hat{y}_t^{(i)} \end{pmatrix} + \begin{pmatrix} X_t^{(R)} \\ Y_t^{(R)} \end{pmatrix}$$

と計算される。提案手法では、各時刻においてこの値を計算することで、入力データを生成する。

3.3.3 移動経路の予測

3.2.2 節において構築された予測器に、3.3.2 節で生成されたデータ列（複数歩行者の経路データ）を入力することで、ターゲットとなる歩行者の移動経路を予測する。

4. 実験

本稿では、異なる種類のデータセットを用いて構築した 2 種類の予測器を利用することで、データセットの違いが予測結果に与える影響を評価する。なお、追従ロボットにより取得された実データを用いた経路予測については、今後の課題とする。

4.1 実験の方法

我々が考察対象としている人物追従ロボットは、博物館*3への設置を予定している。そのため、テストデータとして、考察対象と近い環境のデータセットを利用することにした。具体的には、対象となる環境は以下の特徴を有している。

- 時間帯によりばらつきはあるが、混雑した時間帯が存在する。
- 歩行者が 3 人ほど並んで通ることができる、屋内の通路である。
- 歩行者は双方向に移動し、すれ違う可能性が高い。

そのため、前述したデータセットの中から、ETH のテスト用データセットを選択することにした。これは、時間帯によって人通りにばらつきはあるものの、混雑した時間帯が存在し、歩行者同士のすれ違いが多数発生しているデータである。また、屋外ではあるが、双方向に移動可能な通路であり、歩行者が 4~5 人程度並んで歩けるような環境のものである。本稿では、このデータセットをテストデータとして利用することにより、対象環境を模擬する。

予測器構築用の訓練データとしては、対象環境と類似し

*3 具体的には、深川江戸資料館 (<https://www.kcg.or.jp>) への設置を予定している。

表 2 Social LSTM と他手法との比較結果

手法	Average Error	Final Error
Obstacle LSTM	2.11	3.12
Vanilla LSTM	2.11	3.11
Social LSTM	1.39	2.10

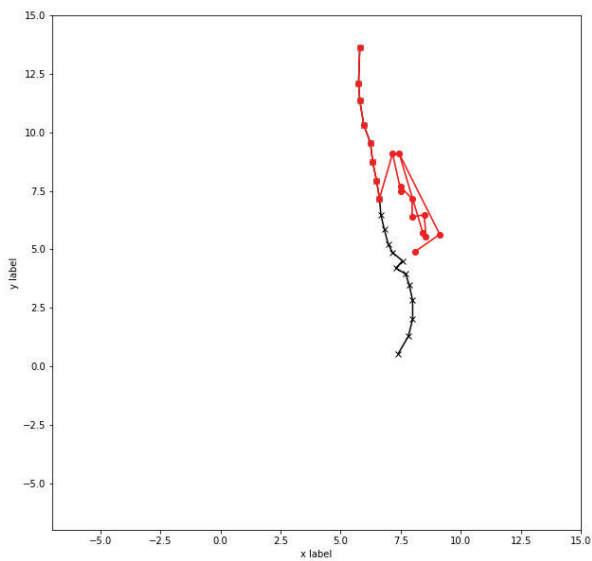


図 4 予測器 1 による予測結果 (赤：予測値，黒：真値)

ているものと異なるものの2種類を利用する。類似しているものとしてはETHの訓練用データセットを、異なるものとしては前述したPETS 2009 Benchmark Dataを利用する。後者は、歩行者が3人ほど並んで通ることができる屋外の通路であるが、人通りが少なく、歩行者同士がすれ違う可能性が低いデータである。

本実験では、これら2つのデータセットを用いて、Social LSTMにより予測器を構築する。ここでは、前者を予測器1 (ETH)、後者を予測器2 (PETS) とする。

4.2 実験の結果

各予測器を用いた実験の結果を表3に示す。評価尺度は、3.2.2項と同様、Average ErrorとFinal Errorである。この結果より、訓練データセットの違いが予測結果に影響を与えていることがわかる。また、移動経路の予測結果を図4と図5に示す。

5. 考察

今回の実験により、ターゲットとなる環境と、予測機を構築するための訓練データを取得する環境との差異が、予測結果に与える影響を調べることができた。実際、表3からは、予測器1の方が、予測器2よりもエラー率が低いことがわかる。この理由は2つ考えられる。一つは、ETHとPETSの環境差が大きい点である。ETHは道幅が広く、

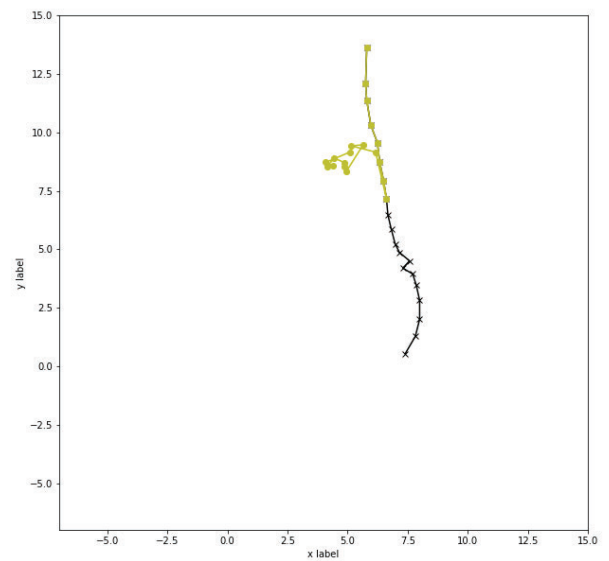


図 5 予測器 2 による予測結果 (黄：予測値，黒：真値)

対向者（歩行者のすれ違い）も多いデータである。一方、PETSは道幅が狭く、人通りも少ない。この差が結果に影響を与えていると考えられる。もう一つは、ETHは観測エリア内に多数の歩行者が存在している点である。そのため、対象となる歩行者と周辺の歩行者との位置情報が、Social LSTMの隠れ層において多数共有されることになった。一方、PETSは観測エリア内の歩行者数がそれほど多くはなく、学習時に、歩行者同士の位置情報はあまり共有されていない。テストデータは歩行者が比較的多くいるデータであるため、すれ違いも多数生じている。そのため、混雑したデータでの学習結果により、人と人の相互作用がモデルに組み込まれた予測器の方が、予測精度が高くなったと考えられる。

これらの結果より、高精度の予測には、適切なデータセットの選択が必要であると考えられる。今回想定する、追従ロボットの利用環境は、博物館等の屋内を想定している。今後は、混雑度合い、地理形状、人の属性、移動速度等、想定環境と似た環境でのデータセットを選択し、検証実験を進めていきたい。今回対象とした、UCY, ETH, PETS以外のデータセットの利用も検討していく。なお、予測アルゴリズムとしては、人と人との相互作用をモデルに組み込むことのできるSocial LSTMの利用は有効であると考えられる。近年、環境側のデータを取り込むことでさらなる精度の向上を目指す研究等も行われているが、本稿での考察対象では、環境側にセンサーを設置することは難しい。適切なデータセットの構築を検討するとともに、人移動モデルと機械学習の統合等も考えていきたい。

表 3 実験の結果

予測器の種類	Average Error	Final Error
予測器 1 (ETH)	2.43	5.78
予測器 2 (PETS)	2.67	7.88

6. おわりに

本稿では、人物追従ロボットの追従精度向上のために、機械学習によりターゲットとなる歩行者の未来の経路を予測する手法を提案した。予測モデルとしては、系列データとして経路予測を行う Social LSTM を利用し、公開されているデータセットを用いて事前に予測器を構築しておくことで、ロボットに搭載されたセンサーのみでの予測を可能にした。また、2種類のデータセットから構築した予測器により、テストデータの予測を行い、データセットの違いが予測結果に与える影響を評価した。

本稿の結論は以下の2点である。

- 追従ロボット用の歩行者経路予測手法として、ロボットに搭載されたセンサーのみで予測を行う手法を提案した。提案手法は、学習済予測器を用いることで、環境側へのセンサー配備を不要にした点、人と人との相互作用をモデルに組み込んだ Social LSTM を利用することにより、混雑環境に対応した点に特徴がある。
- 学習済予測器を構築するためのデータセットは、対象環境と類似したものが望ましいことを検証した。具体的には、環境の地理形状の他、歩行者がすれ違う距離、すれ違い頻度等が対象環境と類似していることが望ましい。

今回は、実際に追従ロボットが取得した実データを用いた経路予測実験は行えなかった。今後、深川江戸博物館等、実環境において、手法の有効性を検証していく予定である。さらに、以下の2点を今後の課題としてあげておく。一つは、使用環境に適した予測手法の選択である。本稿では Social LSTM の利用を前提としたが、経路予測については日々様々な手法が開発されており、環境により適した手法も異なってくると考えられる。これらの比較検証を行う必要がある。二つ目は、使用環境に適したデータセットの選択である。本稿における実験では、データセットの影響について、限られた実験しか行えなかった。今後、適切な訓練データの構成手法を検討していく必要がある。

謝辞 本研究の一部は、JSPS 科研費 17K00366 の助成を受けたものである。

参考文献

- [1] Shimoyama, M., Matsuhira, N. and Suzuki, K.: Human Characterization by a Following Robot Using a Depth Sensor, *Proc. IEEE/SICE International Symposium on System Integration (SII 2017)*, pp. 998–1003 (2017).
- [2] Helbing, D. and Molnar, P.: Social force model for pedestrian dynamics, *Physical Review E*, Vol. 51, No. 5, p. 4282 (1995).
- [3] Alahi, A., Goel, K., Ramanathan, V., Robicquet, A., Fei-Fei, L. and Savarese, S.: Social LSTM: Human Tra-

- jectory Prediction in Crowded Spaces, *Proc. IEEE Conference on Computer Vision and Pattern Recognition (CVPR 2016)*, pp. 961–971 (2016).
- [4] Gupta, A., Johnson, J., Fei-Fei, L., Savarese, S. and Alahi, A.: Social GAN: Socially Acceptable Trajectories With Generative Adversarial Networks, *Proc. IEEE Conference on Computer Vision and Pattern Recognition (CVPR 2018)*, pp. 2255–2264 (2018).
- [5] 加藤由花, 池田貴政, 岡野 憲, 松日楽信人: インタフェースロボット応答制御のための歩行者分岐方向の予測, 情報処理学会論文誌, Vol. 60, No. 2, pp. 572–580 (2019).
- [6] 平川 翼, 山下隆義, 玉木 徹, 藤吉弘亘: 動画を用了経路予測手法の分類, 電子情報通信学会論文誌, Vol. J102-D, No. 2, pp. 53–67 (2019).
- [7] Pellegrini, S., Ess, A., Schindler, K. and van Gool, L.: You'll Never Walk Alone: Modeling Social Behavior for Multi-target Tracking, *Proc. IEEE International Conference on Computer Vision (ICCV 2009)*, pp. 261–268 (2009).
- [8] Hochreiter, S. and Schmidhuber, J.: Long Short-Term Memory, *Neural Computation*, Vol. 9, No. 8, pp. 1735–1780 (1997).
- [9] Sadeghian, A., Alahi, A. and Savarese, S.: Tracking the Untrackable: Learning to Track Multiple Cues with Long-Term Dependencies, *Proc. IEEE International Conference on Computer Vision (ICCV 2017)*, pp. 300–311 (2017).
- [10] Xiang, Y., Alahi, A. and Savarese, S.: Learning to Track: Online Multi-Object Tracking by Decision Making, *Proc. IEEE International Conference on Computer Vision (ICCV 2015)*, pp. 4705–4713 (2015).
- [11] Chen, L., Ai, H., Zhuang, Z. and Shang, C.: Real-Time Multiple People Tracking with Deeply Learned Candidate Selection and Person Re-Identification, *Proc. IEEE International Conference on Multimedia and Expo (ICME 2018)*, pp. 1–6 (2018).
- [12] 奥迫伸一, 坂根茂幸: レーザレンジファインダを用いた移動ロボットによる人の追跡, 日本ロボット学会誌, Vol. 24, No. 5, pp. 605–613 (2006).
- [13] Lerner, A., Chrysanthou, Y. and Lischinski, D.: Crowds by Example, *Computer Graphics Forum*, Vol. 26, No. 3, pp. 655–664 (2007).
- [14] 坂井 栞, 木村純麗, 池田貴政, 野見山大基, 松日楽信人, 加藤由花: 測域センサにより取得される歩行パターンを利用した高齢者/若年者弁別手法, 情報処理学会論文誌, Vol. 58, No. 2, pp. 375–383 (2017).