

移動対象の属性と環境情報を導入した LSTM による経路予測

箕浦大晃 平川翼 山下隆義 藤吉弘亘

中部大学

E-mail: himi1208@mprg.cs.chubu.ac.jp

Abstract

経路予測とは、歩行者や自動車などの移動物体が未来にどのような道筋を移動するかを予測する問題である。深層学習における経路予測には、歩行者間の相互作用を考慮した予測手法や予測対象周囲の動的な環境に関する特徴を考慮した予測手法が提案されている。しかし、これらの手法では、全ての予測対象を単一の種類として扱っており、予測対象の種類に応じた経路を予測することが困難であるという問題点がある。そこで本研究では、予測対象の種類に応じた経路予測の実現を目的とし、対象の属性および予測対象周囲の環境情報を導入した経路予測手法を提案する。提案手法では、対象の過去の移動情報、属性および周囲の環境を表現した特徴ベクトルを LSTM に入力する。これにより、対象毎に適した経路の予測が可能となる。評価実験では、属性および環境に関する特徴を入力することで、対象の移動情報のみを入力する場合と比較して、高精度な予測が実現できることを示す。

1 はじめに

経路予測は歩行者や自動車などの予測対象が未来にどのような経路で移動するかを予測する問題である。経路予測は自動運転による交通事故の防止 [10, 12, 14] やロボットの自律制御 [15, 18, 20, 21] などに応用が期待されており、近年盛んに取り組まれている問題の一つである。すでに様々な予測手法が提案されているが、特に近年では、Convolutional Neural Network (CNN) [3] および Long Short-Term Memory (LSTM) [4] を用いた予測手法が提案されている [1, 2, 5, 6, 7, 15]。

経路を正確に予測するために、従来手法では周囲の歩行者の動きや環境などの情報が導入されている。しかしながら、これらの手法では全ての予測対象を同一種類の予測対象として扱っているという問題がある。現実のシーンでは歩行者だけでなく自動車や自転車などの異なる移動対象が存在する環境下で予測を行う必要がある。その際、対象の種類により移動速度や移動する領域が異なることが考えられる。そのようなシーン

において歩行者や自動車といった複数種類の対象の経路を同時に予測する場合、従来の予測手法では同一種類の対象として予測しているため、予測対象の種類に応じた経路を予測することが困難である。上記の問題を解決するためには、対象の種類毎にモデルを作成し、予測を行うことが考えられる。しかし、対象の種類が増加するにつれて、扱うモデルの数が増加するため、計算コストの面から現実的とは言えない。

そこで本研究では、予測対象の種類および対象周囲の環境情報を導入した経路予測手法を提案する。具体的には、対象の属性、対象周囲の環境、過去の対象の位置座標の3つの情報を使用する。対象の属性では、歩行者や自動車といった予測対象の種類を対象が保有する属性と見なし、対象の属性情報を one-hot vector として表現する。環境情報では、セマンティックなシーンラベルを CNN へ入力し、環境に関する特徴ベクトルを抽出する。過去の対象の位置情報では、各時刻における予測対象の座標を抽出する。これらの情報を LSTM に入力することで、ネットワークの出力として次の時刻の位置情報を得る。予測時には、ネットワークの出力を次の時刻の入力として逐次的に入力することで、過去の情報を考慮した予測を実現する。これにより、対象の移動速度の違いや移動する領域の違いを考慮した経路予測を実現する。

さらに、本稿では絶対座標と相対座標の2種類の異なる位置情報の入力による精度比較を行う。従来の予測方法では、予測対象の過去の位置情報を表す入力として絶対座標を用いている。一方、提案手法では相対座標、すなわち連続する2つの座標の差から得られる方向と大きさを使用する。相対座標を導入することで、予測結果が学習したシーンに依存することを防ぎ、異なるシーンでの経路予測を可能にする。

2 関連研究

経路予測は古くから扱われてきた問題であり、現在までに数多くの研究がなされている。代表的なアプローチとしてベイズモデルに基づく手法がある [10, 12]。Schneider ら [10] は車載カメラに映った車両前方の歩行者の移動経路を予測することを目的として、拡張カルマン

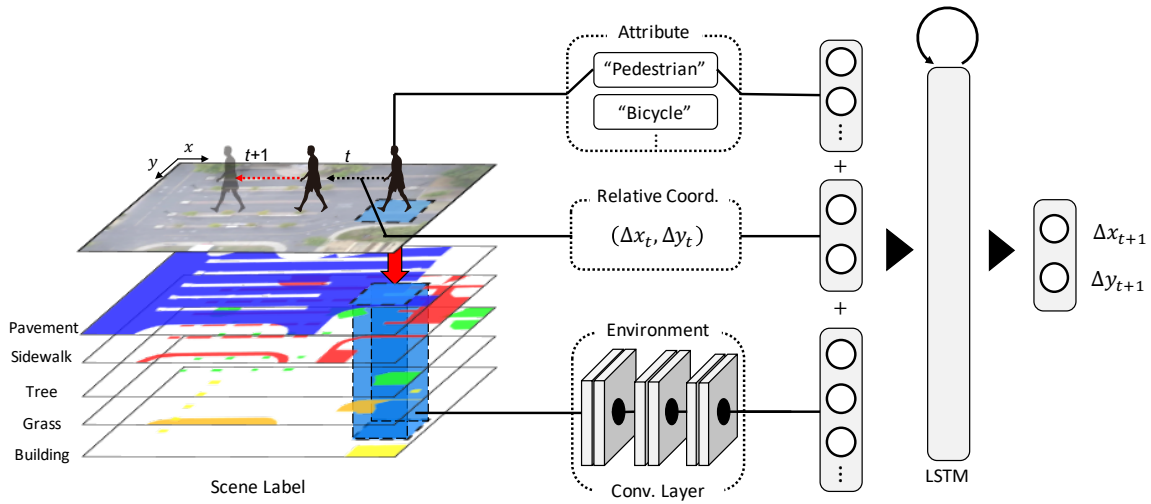


図 1 提案手法のネットワーク構成

フィルタに基づいた経路予測手法を提案している。この手法では歩行者の速度や加速度などの対象の移動に関する情報をいくつか変更することで、予測に有効な情報を検討している。また、Kooij ら [12] は Dynamic Bayesian Network (DBN) [19] を用いて、歩行者が車道を横切る際の移動を予測している。この手法では、歩行者の頭部の向きや車両との距離、縁石までの距離を DBN の観測に使用することで、精度の向上を図っている。一方、提案手法では、対象の種類を対象が保有する情報として導入し、セマンティックな領域ラベルを周囲の情報として導入する。

近年では深層学習、特に LSTM を用いた経路予測手法が提案されている。Alahi ら [1] は、シーン中の複数の歩行者の移動経路を予測する手法を提案している。その際、歩行者同士の衝突を避けるような動きを予測するために、Social Pooling (S-Pooling) と呼ばれるプーリング層を提案している。自身の周辺に存在する歩行者の LSTM の中間層出力を S-Pooling に入力し、その出力を次の時刻の LSTM への入力として用いることで、歩行者同士の衝突を避けるような相互作用を考慮した経路予測を実現している。また、Lee ら [2] は RNN Encoder-Decoder [13] および Conditional Variational Auto-encoder [17] を用いた経路予測手法を提案している。この手法では、対象間の相互作用に加えて、周囲の環境情報を考慮することで高い予測精度を実現している。このように LSTM を用いた予測手法では、対象の周囲の環境に関する特徴ベクトルを記述し、活用することで精度の向上を図っている。一方、提案手法では、周囲の環境に加えて対象の属性情報を導入する。

対象の属性を考慮した手法として、Ma ら [11] は歩行者の性別や年齢を考慮した手法を提案している。この手法では、歩行者の年齢および性別によって歩行速度

表 1 提案手法のネットワーク構成の詳細

層・処理	カーネルサイズ	出力サイズ	備考
input		(100, 100, 7)	
conv1	(5, 5)	(48, 48, 16)	ReLU
norm1		(48, 48, 16)	Batch Norm
pool1	(2, 2)	(24, 24, 16)	MaxPooling
conv2	(5, 5)	(20, 20, 32)	ReLU
norm2		(20, 20, 32)	Batch Norm
pool2	(2, 2)	(10, 10, 32)	MaxPooling
conv3	(5, 5)	(6, 6, 32)	ReLU
pool3	(2, 2)	(3, 3, 32)	MaxPooling
concat		296	
LSTM		128	
output		2	

が異なることを仮定し、シーン中に存在する複数の歩行者の移動経路を同時に予測している。この手法では歩行者のみが存在するシーンで予測を行っており、環境に関する情報は用いられていない。一方、本研究では、対象の属性および周囲の環境情報を考慮した経路予測を行う。

3 提案手法

本節では、予測対象の種類および対象周囲の環境情報を導入した経路予測手法について述べる。提案手法のネットワーク構成を図 1 に、詳細を表 1 に示す。はじめに、予測対象の属性を表現するために、属性を one-hot vector で表す。次に、シーン中の予測対象を中心とする静的なシーンラベルを CNN に入力し、予測対象周囲の環境に関する特徴ベクトルを抽出する。これらのベクトルを過去の対象の動きと連結し、LSTM へ入力する。LSTM の出力として、次の時刻の対象の座標を得る。出力される座標は、入力する座標と同一種類である。以下に各入力情報の表現方法、およびネットワークへの入力について説明する。

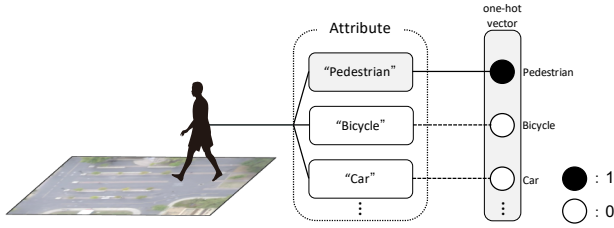


図2 対象の属性の表現方法. 対象の属性が歩行者である場合を示す.

3.1 属性情報

複数種類の対象の経路を予測するためには、対象の属性を表す情報を導入する必要がある。属性情報をネットワークへ入力するために、本研究では対象の属性情報を one-hot vector で表現する。図2に属性情報の表現方法を示す。具体的には、属性の数を N_a とし、 N_a 次元のベクトルを用意する。そして、対象の種類に対応するベクトルの要素を1、それ以外の要素を0とすることで、予測対象の属性を表現する。予測対象の属性情報を導入することで、対象の種類に応じた移動速度を考慮した予測が期待できる。また、環境情報との組み合わせにより対象が移動しやすい領域を考慮することが期待できる。

3.2 移動情報

ベイズモデルや LSTM を用いた従来の経路予測手法では、各時刻において予測対象が存在する位置（絶対座標）を用いて予測が行われている。しかしながら、絶対座標を用いた経路予測では、複数のシーンで予測を行うことが困難であるという問題が存在する。複数のシーンの予測を行う場合、予測シーン毎に建物や木などの障害物、歩道や車道などの移動領域が異なる。絶対座標を用いた経路予測では、このような領域の情報が暗に考慮されるため、複数のシーンでの予測が困難となる。

そこで本研究では、移動量、すなわち前後の時刻間での相対的な座標を用いた予測を行う。具体的には、過去から現在の位置情報、すなわち絶対座標の差分から相対座標 $(\Delta x_t, \Delta y_t)$ を求める。求めた相対座標を LSTM へ入力することで、次の時刻における相対座標を得る。相対座標を用いることで、対象の現在地を常に基点とすることができ、建物などの障害物や歩道などの移動領域の位置情報に依存することなく、複数のシーンでの経路の予測を行うことが可能になると考えられる。実験にて、絶対座標および相対座標を入力とした場合の予測精度の変化を検証する。

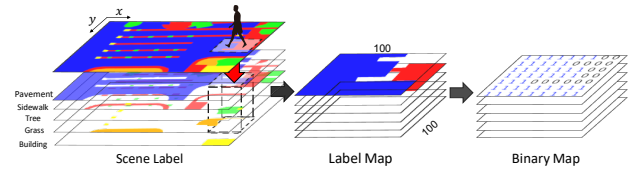


図3 環境に関する特徴ベクトルの抽出

3.3 環境情報

予測対象周囲の環境情報を表現するために、本研究では、歩道や建物などのシーンに付与されたセマンティックなシーンラベルを用いて環境情報の抽出を行う。環境情報の導入手順を図3に示す。まず、シーンラベルから予測対象を中心とする 100×100 [pixel] の領域のラベルを抽出する。次に、ラベルの種類毎に分割したラベルマップを作成する。そして、ラベルマップを0または1で表現されたバイナリマップに変換する。作成したバイナリマップを CNN へ入力することで、環境に関する特徴量を抽出する。環境情報を導入することで、障害物の有無や対象の属性に応じた領域を考慮する経路予測が可能になる。

3.4 ネットワークへの入力方法

上記の属性、移動情報、および環境情報を LSTM へ入力することで次の時刻における対象の位置を予測する。具体的には、予測対象が実際に移動した真値および、その座標を中心として抽出したシーンラベル、対象の属性を観測データとして用いる。LSTM にはこれらの観測データを予測開始直前のフレームまで逐次的に入力する。予測時には LSTM の出力である予測値を次の時刻の入力として逐次的に入力する。その処理を予測終了時刻まで行い、予測の実現をする。

4 評価実験

本節では、提案手法の有効性を評価実験により検証する。

4.1 データセット

データセットには、Stanford Drone Dataset (SDD) [8] を用いる。SDD には bookstore, coupa などの8つの異なる予測シーンから構成される。各シーンには異なる日時で撮影された動画数が複数含まれており、合計で60本の動画像から構成される。また、SDD は biker, pedestrian, cart, car, bus, skateboarder の異なる6種類の移動対象と座標データから構成されている。本実験では、6種類の異なる属性を使用する。また、SDD はフレームレート 30 fps で撮影されており、本実験では20フレーム毎の座標データを用いて実験を行う。すなわち、予測の1ステップは約 0.66 [s]

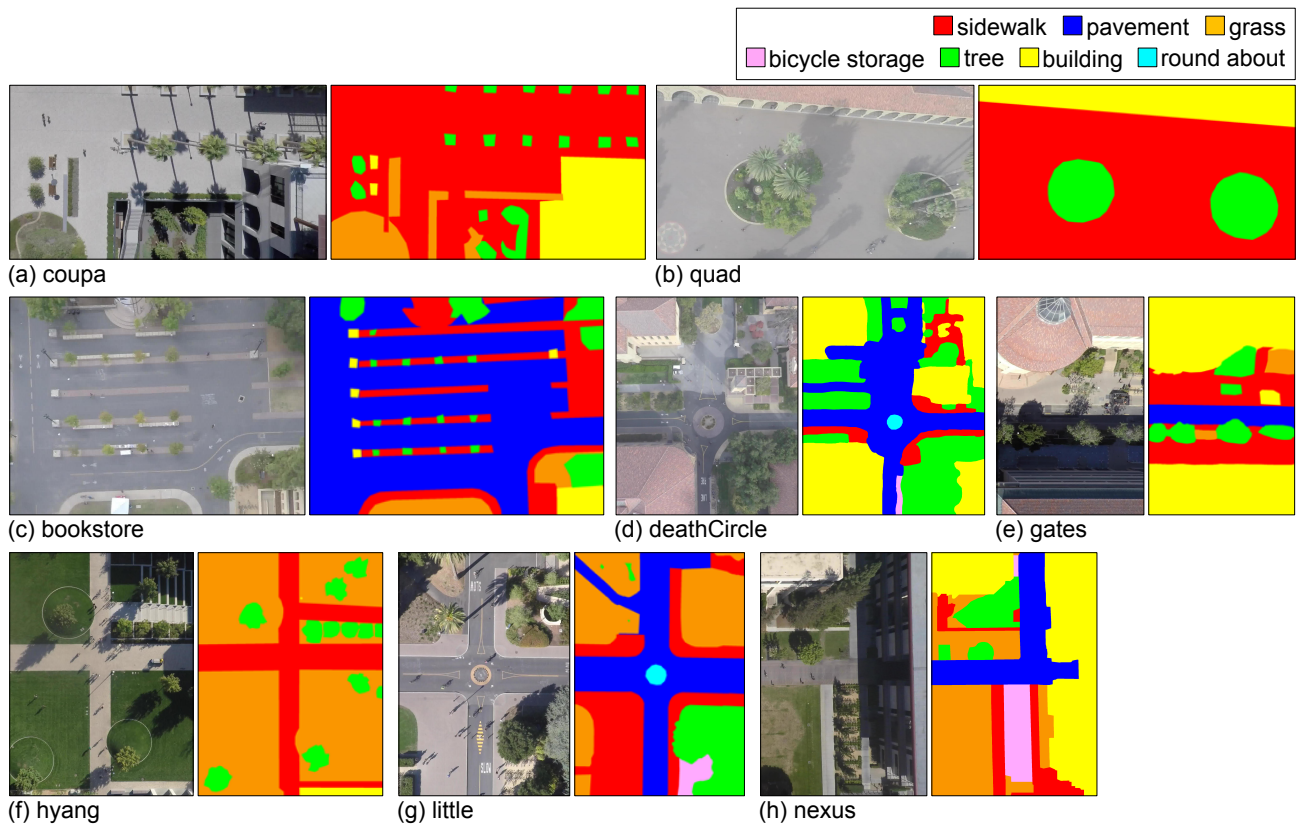


図4 SDDに付与したシーンラベルの例. 左図は実際のシーン画像, 右図は対応するシーンラベルを示す.

に対応する. 本実験では, 5ステップ(約3.3秒)を観測として使用し, その後の8ステップ(約5.3秒)を予測する.

提案手法では, 環境情報を表現するために, シーンラベルを入力として用いる. しかし, SDDにはシーンラベルの情報は含まれていないため, 全てのシーンに対してラベルを付与した. 付与したラベルの種類はsidewalk, pavement, grass, bicycle storage, tree, building, round aboutの計7種類である. 図4にシーンラベルの例を示す. 予測対象の数は学習用に5,365, 評価用に1,082を使用する. 使用するデータの内訳を表2に示す.

4.2 評価指標

定量評価のために, 本実験では2種類の評価指標を用いる. 1つ目は予測の最終フレームにおける真値と予測値のユークリッド距離であるFinal displacement error. 2つ目は各予測フレームにおける真値と予測値のユークリッド距離の平均であるAverage displacement errorを用いる.

また, 比較手法として, Kalman Filter (KF) [16]を用いる.

4.3 学習条件

学習条件としてバッチサイズを10に設定し, エポックを100とした. また, 最適化手法にRMSprop [9]を用いた. 学習時には, 観測の開始から予測終了までの

表2 使用するデータの内訳

		train	test
# of scenes		52	8
attribute	biker	2,369	545
	pedestrian	2,696	500
	cart	71	15
	car	75	5
	bus	17	2
	skateboarder	137	15

全ステップ間, 対象の移動情報として真値を入力する. 深層学習のフレームワークとしてChainerを使用し, NVIDIA Geforce 1080を用いてend-to-endで学習を行った.

4.4 相対座標を入力とした場合の実験結果

相対座標を使用した場合の予測精度の定量的評価を表3, 予測結果例を図5に示す. KFでは過去の移動情報のみを観測として使用しており, 線形な移動経路を予測している. また, 障害物などの周囲の環境情報を用いていないため, 図5(b)に示すような障害物へ衝突するような予測を行なっている.

表3より, 相対座標に加え, 属性情報または環境情報をLSTMに導入することで, 予測精度が向上していることがわかる. 特に環境情報を導入した場合(相対座標+環境)では, 図5(h)に示すような障害物を避ける予測が得られる. また, 提案手法である相対座標+属性

表 3 相対座標を入力とする場合の評価結果 (単位: pixel)

入力	Final disp. error	Avg. disp. error
KF	174	116
相対座標	196	86
相対座標+属性	173	76
相対座標+環境	172	76
相対座標+属性+環境	109	53

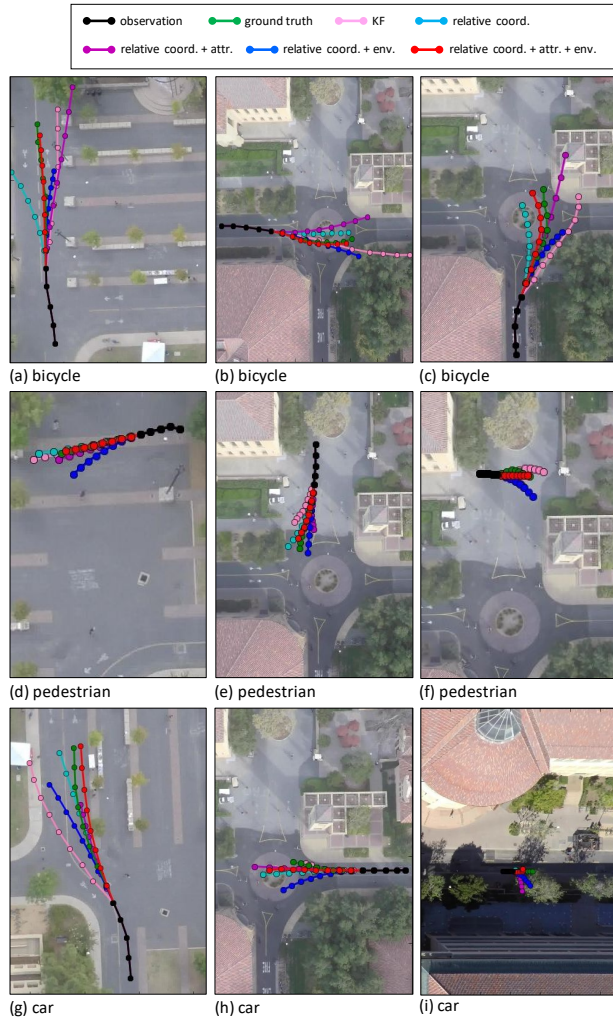


図 5 相対座標を用いた場合の予測結果

+ 環境は他の手法に比べ、予測精度が向上していることがわかる。予測結果を確認した場合でも、真値に近い経路を予測していることが確認できる (図 5(a)(b)(c)(g)).

4.5 絶対座標を入力とした場合の実験結果

本節では移動に関する情報として、絶対座標を用いた場合の実験結果について示す。絶対座標を使用した場合の予測精度の定量的評価を表 4, 予測結果例を図 6 に示す。表 3 および表 4 より、相対座標を入力とした場合が絶対座標を入力とした場合と比較して、全体的に予測精度が向上していることが確認できる。この結果より、LSTM に基づく経路予測には相対座標を用いると大幅な精度の向上が得られることが確認できる。ま

表 4 絶対座標を入力とする場合の評価結果 (単位: pixel)

入力	Final disp. error	Avg. disp. error
KF	142	93
絶対座標	641	305
絶対座標+属性	250	135
絶対座標+環境	293	136
絶対座標+属性+環境	187	102

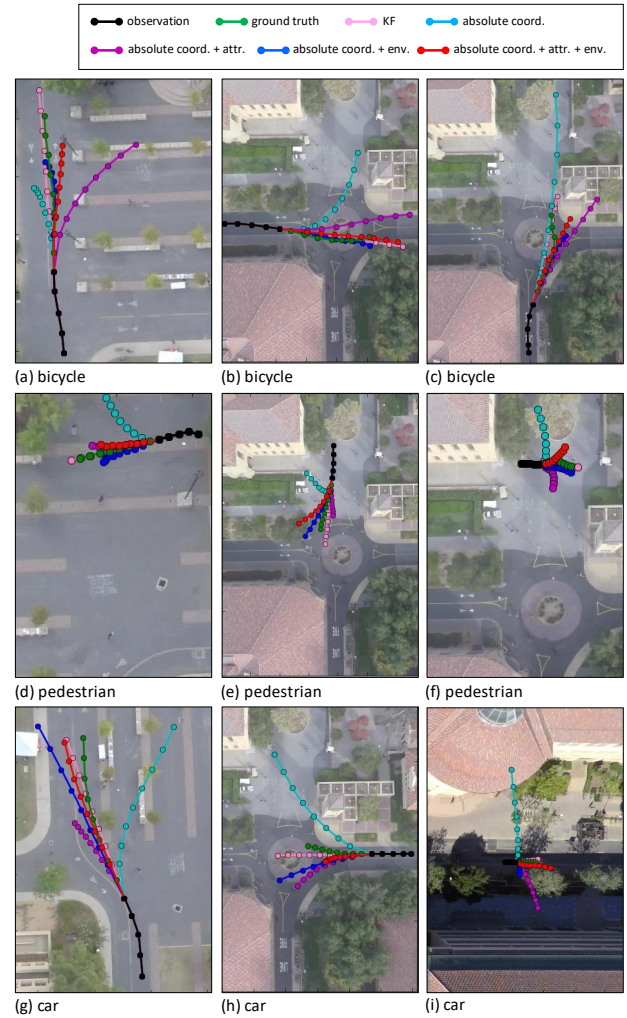


図 6 絶対座標を用いた場合の予測結果

た、絶対座標を用いた LSTM に基づく経路予測手法は絶対座標を用いた KF と比較して精度が低下している。これは、KF が観測された経路から線形に予測を行い、対象が直進する場合に対して頑健な予測結果を得ることが可能なためだと考えられる。しかしながら、KF では図 6(b)(c)(h) のような障害物を考慮した経路の予測は困難である。

絶対座標のみを用いた LSTM に基づく予測手法では、予測結果が真値と大きく異なっていることがわかる。図 5(e) では、対象の移動方向は真値と逆向きの予測をしている。絶対座標に加えて、属性や環境情報を導入することでこの問題は改善されている。しかしな

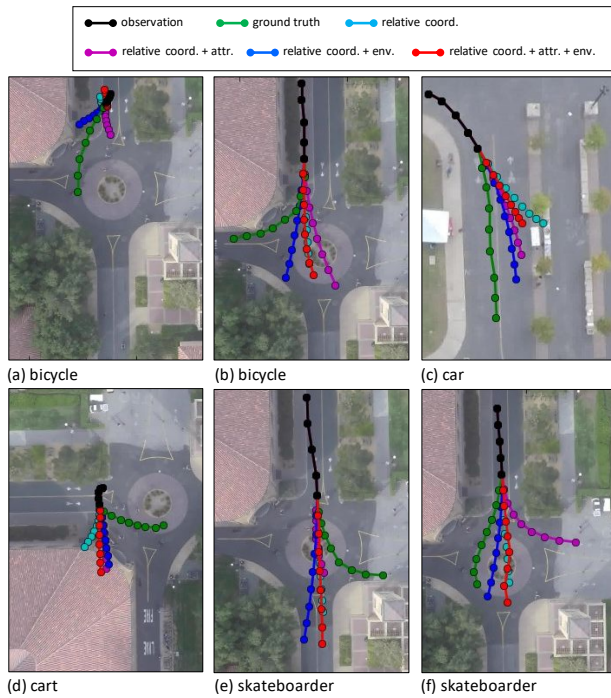


図 7 予測結果の失敗例

がら、相対座標の結果と比較して絶対座標では真値と異なる予測をしていることから、複数シーンにおける LSTM を用いた経路予測では、相対座標を用いることが望ましいと言える。

4.6 予測の失敗例

図 7 に、適切な予測が行えなかった例を示す。移動情報として相対座標を用いた場合の予測結果を示している。図 7(a) より、急激に移動速度が変化する場合では、その速度変化に対応することができず真値と異なる動きを予測している。また、図 7(b) では観測データが直進しているため予測結果も直進した動きを予測している。しかし、実際には対象は右折した移動をしている。このことから、分岐した経路の予測は困難であると考えられる。図 7(d) では、真値は左折しているが、予測結果は建物方向に直進していることがわかる。図 7(d, e, f) は、cart, および skateboarder の予測結果である。表 2 に示すように、これらの属性に対応するサンプル数は他の属性に比べて著しく少ないために、正しく学習が行えなかったと考えられる。その結果、誤った経路を予測し、障害物を回避できなかったと考えられる。したがって、データ数が少ない属性の場合に対する効率的な学習方法の検討が必要だと考えられる。

5 おわりに

本稿では、予測対象の属性、および周囲の環境情報を導入した経路予測手法を提案した。提案手法では、one-hot vector で表現した対象の属性情報とシーンラベル

を CNN に入力することで抽出した環境情報を入力として用いた。また、予測対象の過去の動きでは相対座標を用いた。これらの情報を LSTM へ逐次的に入力することで、予測を実現した。SDD を用いた評価実験により、位置情報のみを入力する場合と比較して属性および環境情報を導入した場合における予測精度が高い結果となった。また、相対座標を用いた場合、絶対座標より予測精度が大幅に向上し、複数シーンでの経路を予測を可能とした。今後の課題として、予測対象間の相互作用、および動的な環境を考慮することが挙げられる。

参考文献

- [1] A. Alahi, K. Goel, V. Ramanathan, A. Robicquet, L. Fei-Fei and S. Savarese, "Social LSTM: Human trajectory prediction in crowded spaces," CVPR, pp. 961-971, 2016.
- [2] N. Lee, W. Choi, P. Vernaza, C.B. Choy, P. H. S. Torr and M.K. Chandraker, "DESIRE: Distant future prediction in dynamic scenes with interacting agents," CVPR, pp. 336-345, 2017.
- [3] Y. Lecun, B. Boser, J.S. Denker, D. Henderson, R.E. Howard, W. Hubbard and L.D. Jackel, "Backpropagation Applied to Handwritten Zip Code Recognition," Neural Computation, pp.541-551, 1989.
- [4] S.Hochreiter, "LONG SHORT-TERM MEMORY," Neural Computation, pp. 1735-1780, 1997.
- [5] T. Fernando, S. Denman, A. McFadyen, S. Sridharan, and C. Fookes, "Tree Memory Networks for Modelling Long-term Temporal Dependencies," arXiv preprint arXiv:1703.04706, 2017.
- [6] T. Fernando, S. Denman, S. Sridharan, and C. Fookes, "Soft + Hardwired Attention: An LSTM Framework for Human Trajectory Prediction and Abnormal Event Detection," arXiv preprint arXiv:1702.05552, 2017.
- [7] S. Yi, H. Li, and X. Wang, "Pedestrian behavior understanding and prediction with deep neural networks," ECCV, pp.263-279, 2016.
- [8] A. Robicquet, A. Sadeghian, A. Alahi, and S. Savarese, "Learning Social Etiquette: Human Trajectory Understanding in Crowded Scenes," ECCV, pp. 549-565, 2016.
- [9] T. Tieleman and G. Hinton, "Lecture 6.5-rmsprop: Divide the gradient by a running average of its recent magnitude," COURSE: Neural networks for machine learning, 2012.
- [10] N. Schneider and D.M. Gavrila, "Pedestrian Path Prediction with Recursive Bayesian Filters: A Comparative Study," GCPR, pp. 174-183, 2013.
- [11] W.C. Ma, D.A. Huang, N. Lee and K.M. Kitani, "Forecasting Interactive Dynamics of Pedestrians with Fictitious Play," CVPR, pp. 774-782, 2017.
- [12] J.F.P. Kooij, N. Schneider, F. Flohr, D.M. Gavrila, "Context-Based Pedestrian Path Prediction," ECCV, pp. 618-633, 2014.
- [13] K. Cho, B. van Merriënboer, C. Gulcehre, F. Bougares, H. Schwenk, and Y. Bengio. "Learning phrase representations using RNN encoder-decoder for statistical machine translation," arXiv preprint arXiv:1406.1078, 2014.

- [14] C.G. Keller and D.M. Gavrilu, "Will the pedestrian cross? a study on pedestrian path prediction," IEEE Transactions on Intelligent Transportation Systems, vol.15, no.2, pp.494-506, 2014.
- [15] A. Vemula, K. Muelling and J. Oh, "Attention: Modeling Attention in Human Crowds," arXiv preprint, 2017.
- [16] R. E. Kalman, "A new approach to linear filtering and prediction problems," J. basic Eng., vol.82, no.1, pp.35-45, 1960.
- [17] D. P. Kingma, S. Mohamed, D. Jimenez Rezende, and M. Welling. "Semi-supervised learning with deep generative models," NIPS, pp. 3581-3589, 2014
- [18] V. Karasev, A. Ayvaci, B. Heisele, and S. Soatto, "Intent-aware, long-term prediction of pedestrian motion," ICRA, pp. 2543-2549, 2016.
- [19] J. W. Robinson and A. J. Hartemink. "Non-stationary dynamic bayesian networks," NIPS, pp. 1369-1376, 2009.
- [20] A. Vemula, K. Muelling and J. Oh, "Modeling cooperative navigation in dense human crowds," ICRA, pp. 1685-1692, 2017.
- [21] B.D. Ziebart, N. Ratliff, G. Gallagher, C. Mertz, K. Peterson, J.A. Bagnell, M. Hebert, A.K. Dey, and S. Srinivasa, "Planning-based prediction for pedestrians," ICIRS, pp.3931-3936, 2009.