H-013

グラフ理論による歩行者と環境の時空間的依存関係に基づく歩行軌道予測

Trajectory prediction based on spatiotemporal interdependence among pedestrians and context with graph theory

對馬 祐太† 杉本 千佳‡
Yuta Tsushima Chika Sugimoto

1. はじめに

歩行者の移動経路を予測する機械学習モデルは、ITS や自律ロボットなど応用先の広さから盛んに研究が行われている。しかし、先行研究の多くは汎用的な環境データを評価対象としており、混雑した環境下で十分な精度を得られるのかは不明である。また、いずれの研究も歩行者自体や歩行者同士の特徴量取得に焦点を当てているが、環境からの情報取得には単純な CNN を用いているのみであり、環境情報を効果的に活かせているとは言えない。

本研究では、環境中で移動経路に影響を与えうる壁や障害物の境界情報のみで精度の良い軌道予測が可能という仮説の元、歩行者と境界情報をグラフのノードとして捉え時空間依存関係を抽出することで、屋内の混雑環境における高精度な軌道予測モデルの構築を目指す。

2. 先行研究

人物軌道予測を行う際の環境情報の取得手段は、主に一人称視点でのカメラやセンサを用いるものと俯瞰映像を用いるものの 2 つに大別される。俯瞰映像を用いた人物軌道予測では、Attention 機構を用いて歩行者同士の相互インタラクションを学習し、複数経路を予測する手法が主流となっている[1]。Zhang らが提案した SR-LSTM[2]では、他歩行者との衝突を防ぐ Pedestrian-aware Attention (PA)と他歩行者の動きを加味した経路選択機構 Motion gate (MG)の 2 つを組み合わせることで、対象人物の近隣に着目して経路を予測した。2020 年に C. Yu らによって発表された STAR[3]は、自然言語処理分野で目覚ましい成果をあげている Transformer[4]を応用した軌道予測モデルである。LSTM や RNN を用いず完全に Attention 機構のみで軌道予測を実現しており、Temporal Transformer ブロックで各歩行者それぞれの時間的依存性を、Spatial Transformer ブロックで各歩行者間の空間的依存性を抽出している。C. Yu らの最大の貢献は、Self-Attention 機構が無向完全連結グラフ上でのメッセージパッシングとみなすことができるとして、Transformer ベースのグラフ畳み込み TGConv を提案したことである。他にも Attention ベースの軌道予測手法は数多く提案されており、いずれも精度評価に用いられているデータセットは ETH Dataset[5]および UCY Dataset[6]である。これらは市街地や大学構内の俯瞰映像から作成されたデータセットで、閑散とした環境も多く含まれたデータに対し学習や評価を行っており、混雑した環境で十分な精度を得られるのかは不明である。また、障害物などの環境情報を正確に取得するための工夫は必ずしもされていない。これらの現状を踏まえ、歩行者間の相互インタラクションを十分抽出できているモデルならば、環境情報を取り込む際に機械学習を用いる必要性は低いのではないかと考えた。人物軌道予測タスクでは、評価方法として 8 フレームの観測データを元にその後の 12 フレームの軌道を予測するのが通例となっている[1]。この評価方法は遠い未来を予測するわけではないため、撮影範囲が狭いシーンを除いて柱や壁といった意味情報を使う必要性は低いのではないかと考えられる。また CNN は画像中の色の影響を大きく受けるため、実世界では太陽光や照明の影響が大きいことを考慮すると、必ずしも適切な学習手法とは言えない。そこで、本研究では混雑環境下でも高精度な予測が可能なモデルに、歩行が可能な場所かどうかを表す境界情報を、意味を持たないノードとして与える手法の性能を評価する。

† 横浜国立大学大学院理工学府 Graduate School of Engineering Science, Yokohama National University
‡ 横浜国立大学大学院工学研究院 Faculty of Engineering, Yokohama National University

3. 提案手法

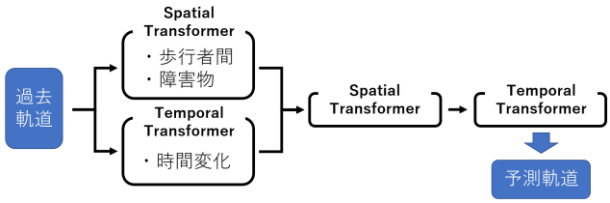

図 1 提案手法のモデル概要図

図 1 に示すように、混雑環境下でも有用な STAR[3]モデルをベースに、静的環境情報をノードとして学習に取り込む。一般的に Attention は、$embeding\{h_t\}_{t=1}^{T}$に対してクエリ行列$Q = f_Q(\{h_t\}_{t=1}^{T})$、キー行列$K = f_K(\{h_t\}_{t=1}^{T})$、バリュー行列$V = f_V(\{h_t\}_{t=1}^{T})$を用いて式(3.1)のようになる。

$$Att(Q,K,V) = \frac{Softmax(QK^T)}{\sqrt{d_k}}V \tag{3.1}$$

なお、d_kは各クエリの次元を表しており、学習の安定のために用いられている。これに対し、グラフ上では特徴集合$\{h_i\}_{i=1}^{n}$の特徴ベクトルh_iに対して、対応するクエリ行列$q_i = f_Q(h_i)$、キー行列$k_i = f_K(h_i)$、バリュー行列$v_i = f_V(h_i)$と表現できる。ノードiからノードjへのメッセージを$m^{i\to j} = q_i^T k_j$と定義すると、式(3.1)を次のように書き換える事ができる。

$$Att(Q,K,V) = \frac{Softmax\left(\left[m^{j\to i}\right]_{i,j=1:n}\right)}{\sqrt{d_k}}[v_i]_{i=1}^{n} \tag{3.2}$$

上記に基づいて提案されたグラフ畳み込み演算 TGConv により、歩行者間の相互インタラクションを適切に抽出することができるため、STAR モデルは混雑環境下でも精度を低下させることなく予測が可能である。本研究ではこの Spatial Transformer ブロックに着目し、各歩行者間の相互インタラクションだけでなく、壁や障害物もノードと捉えることで環境情報の予測への取り込みを可能にした。実世界の屋内環境は人工的に作られたものであるため、壁や障害物は直線的であるシーンも多い。アノテーションの際にはまずシーン画像から通路の曲がり角部分や折れ点となる箇所のピクセル位置を取得し、算出した 1 次関数の式を用いてノードの設定を行った。その後、有機的な形状の通路や障害物について手動でアノテーションを行う事で、歩行可能な場所の境界を設定した。この手法では各ノードが障害物や壁といったラベル情報を持たず、全て単純な独立したノードとしてモデルに与えている。これにより、予測時の計算量の削減、アノテーションの効率化などを達成した。提案モデルは各歩行者との相互インタラクションを学習することで、壁や障害物を考慮した軌道を予測することが可能になる。

4. 実験

本章では、提案手法による人物軌道予測の精度を評価する。

4.1 実験諸元

本実験では、Grand Central Station Dataset (GC) [7,8]および ATC Dataset[9]の 2 つを用いる。いずれも屋内の混雑した環境下で取得されたデータであり、各タイムフレームごとの歩行者の ID、座標が CSV 形式で与えられている。実験時にはデータをそれぞれ 4 つに分割し、3 つ目のデータをテストデータとし、残りを訓練データとした。環境情報のノードは、国土交通省道路局のガイドラインにより歩行者の歩道幅員の占有幅が 0.75m と規定されている[10]事から、ノード同士が 0.5m 間隔となるように設定した。

評価指標には先行研究で多く用いられている平均変位誤差（ADE）および最終変位誤差（FDE）を用いる。観測フレーム数を 8 フレーム、予測フレーム数を 12 フレームとし、その他の設定については STAR と同一とした。

4.2 実験結果

実験結果を表 2 に示す。いずれのデータセットでの実験においても ADE、FDE 両指標ともに予測精度が向上していることが確認できる。また、予測結果を可視化した例を図 2 に示す。それぞれの軌跡の色は「緑：観測軌跡　赤：予測結果　青：グランドトゥルース」に対応している。図 2 を見ると、障害物を避ける軌跡や狭い通路に沿った軌跡が精度良く予測できていることが確認できる。特に図 2 下段の予測においては、CNN などによるシーンの意味情報を用いなくとも通路に沿った予測が可能となっており、適切に壁や障害物の境界情報を取得できれば、複雑なアルゴリズムを用いなくとも静的な環境情報を考慮することが可能であると考えられる。

表 2　精度結果比較

		GC	ATC
STAR [9]	ADE	0.293	0.928
	FDE	0.939	2.576
Proposed method	ADE	0.270	0.875
	FDE	0.831	2.340
Error decrease rate	ADE	**-7.86%**	**-5.65%**
	FDE	**-11.42%**	**-9.17%**

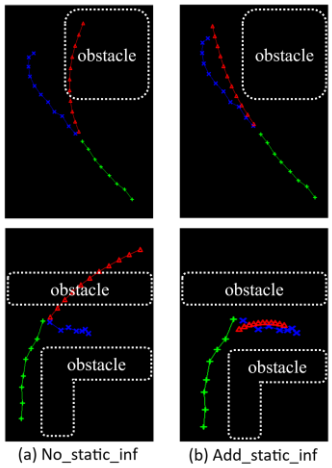

図 2　評価結果の軌跡例

5. おわりに

本研究では、軌道予測モデル STAR を元に意味ラベルを持たない形で環境情報を用いる手法を提案した。先行研究ではシーンの環境情報を CNN などの機械学習によって取り込んでいたが、単純なノードとしてモデルに与えるだけでも予測精度が約6~11%向上することが確認できた。近年の軌道予測研究は特徴量抽出のためにモデル・データの複雑さや求められる計算量が増大しつつあるが、本研究では環境情報の意味ラベル等を用いなくとも実用に耐えうる予測モデルが作成できる可能性を示した。今後の課題として、歩行可能な場所の境界をシーン画像から正確に自動抽出するアルゴリズムを検討することが挙げられる。

参考文献

[1] 箕浦大晃, et al., “ Deep Learning を用いた経路予測の研究動向”, 信学技報, vol.120, no.187, PRMU2020-29, pp.62-78 (2020).
[2] P. Zhang, et al., “SR-LSTM: State Refinement for LSTM Towards Pedestrian Trajectory Prediction”, CVPR, pp.12077-12086 (2019).
[3] C. Yu, et al., “Spatio-Temporal Graph Transformer Networks for Pedestrian Trajectory Prediction”, ECCV (2020).
[4] A. Vaswani, et al., “Attention Is All You Need”, Advances in Neural Information Processing Systems, vol.30 (2017).
[5] S. Pellegrini, et al., “You’ll never walk alone: Modeling social behavior for multi-target tracking”, ICCV, pp.261-268 (2009).
[6] A. Lerner, et al., “Crowds by example. In Computer graphics forum”, Willey Online Library, vol.26, pp.655-664 (2007).
[7] B.Zhou, et al., “Understanding Collective Crowd Behaviors: Learning a Mixture Model of Dynamic Pedestrian-Agents”, CVPR (2012).
[8] S. Yi, et al., “Understanding Pedestrian Behaviors from Stationary Crowd Groups”, CVPR (2015).
[9] D. Brscic, et al., “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).
[10] 国土交通省道路局, “道路の移動等円滑化に関するガイドライン 令和 4 年 6 月”, p.1-1