A Survey on Trajectory Encoding Methods for Social Robots

LEANDRO DE SOUZA ROSA and LUCA IOCCHI*, La Sapienza University of Roma, Italy
JULIAN F. SCHUMANN, Delft University of Technology, Netherlands
ANNA MÉSZÁROS, Delft University of Technology, Netherlands

ACM Reference Format:
Leandro de Souza Rosa, Luca Iocchi, Julian F. Schumann, and Anna Mészáros. 2018. A Survey on Trajectory Encoding Methods for Social Robots. In . ACM, New York, NY, USA, 3 pages. https://doi.org/XXXXXXX.XXXXXXX

1 ABSTRACT

2 INTRODUCTION

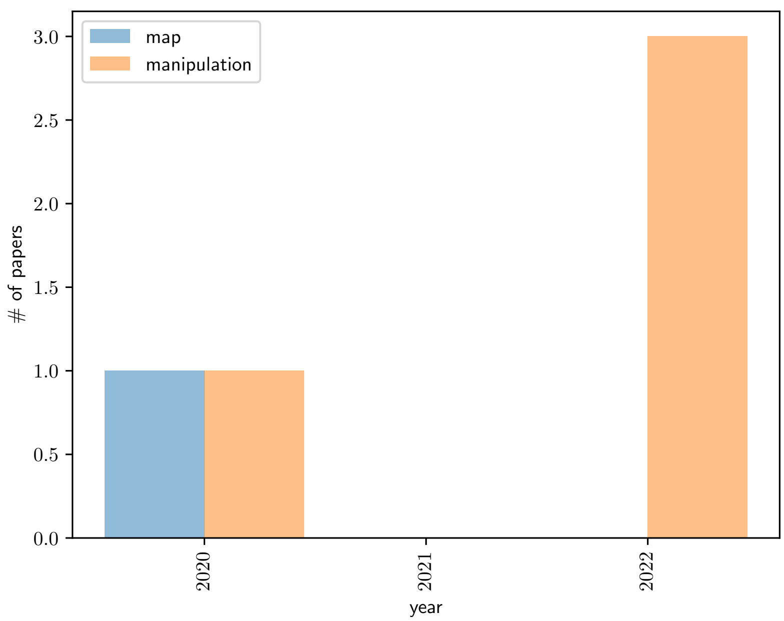

Fig. 1. Number of publications according to Scope.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than the author(s) must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.
© 2018 Copyright held by the owner/author(s). Publication rights licensed to ACM.
Manuscript submitted to ACM

3 RELATED WORKS

4 ENCODINGS

4.1 Recurrent Neural Networks

[just for testing] [5] The paper uses relatively normal combination of rnn and attention to encode the past, its most important contribution comes in the recursive reasoning of the decoder.

4.2 Long-Short Term Memory

[just for testing] [4] focus on learning human-compliant behaviors for robots navigating in human crowded environments, by optimizing trajectories both for comfort (the absence of annoyance and stress for humans) and naturalness (the similarity between robots and humans).

They state "reinforcement learning approaches tend to optimize on the comfort aspect of the socially compliant navigation, whereas the inverse reinforcement learning approaches are designed to achieve natural behavior."

A Long Short-Term Memory (LSTM) encoder-decoder is defined for each of three social-interaction force (intention, social interaction and fluctuation) from [2] to improves interpretability, which are used with a adversarial training to reduce the data-bias tendency of LSTMs.

The trajectories of the robot and other agents (humans) are encoded with LSTMs, for intention just the robot trajectory is encoded, for social interaction and fluctuation all trajectories are encoded and passed through a `maxpooling` layer before the LSTM. Both encoding are passed to an adversarial training using demonstrations.

4.3 Graph Neural Networks

[just for testing] [1] In this paper, they suggest PAGA, which is an improvement on LaneCGNB [3]. Namely, the update the method for calculating the attention weights in the GNN used to encode the map. Instead of simply using an the four different adjacency matrices, the replaces that with a more complex matrix able to include connection via intermediate steps.

5 MULTI-LAYERS

6 MULTI-AGENTS

6.1 1x1

6.2 1xN

6.3 nxN

[just for testing] [6] In this paper, the encoding will be done in multiple steps.

- Extract for each agent its surrounding region (other agents and lanes) and encoded the corresponding past observations, using a coordinate system that is normalized to this agent last observed position and velocity, where we embed the agents own trajectory as well as the surrounding agents trajectory.
- We apply multi head cross attention with the agent as the query and its neighbors as key and values, and a recursive layer afterwards.
- The spatial encodings are then encoded temporally using a time axis transformer block
- Similar to the encoding of the neighbors, the encoded agent position is enhanced by cross attention with the neighboring lane segments (here, each vector in a polyline is treated individually)

- An transformer is used to encode the interactions with each agent, where their states are expanded by the translations and rotations between the coordinate systems.
- The decoder used is simple, however, the joint prediction are not really joint, just parallel

Here, using for the local encoding two different transformers, first along the agent axis and then along the time axis, significantly reduces the computational cost of the model.

6.4 NxN

7 APPLICATIONS

8 HUMAN ASPECTS

9 DATASETS, BENCHMARKS AND SIMULATORS

10 CONCLUSIONS

ACKNOWLEDGMENTS

This project has been sponsored by ...

REFERENCES

[1] Fang Da and Yu Zhang. 2022. Path-aware graph attention for hd maps in motion prediction. In *2022 IEEE International Conference on Robotics and Automation (ICRA)* (Philadelphia, USA). IEEE Computer Society, Los Alamitos, CA, USA, 6430–6436. https://doi.org/10.1109/ICRA46639.2022.9812100

[2] Dirk Helbing and Péter Molnár. 1995. Social force model for pedestrian dynamics. *Phys. Rev. E* 51 (May 1995), 4282–4286. Issue 5. https://doi.org/10.1103/PhysRevE.51.4282

[3] Ming Liang, Bin Yang, Rui Hu, Yun Chen, Renjie Liao, Song Feng, and Raquel Urtasun. 2020. Learning lane graph representations for motion forecasting. In *Computer Vision–ECCV 2020, Proceedings* (Glasgow, UK). Springer, 541–556. https://doi.org/10.1007/978-3-030-58536-5_32

[4] Chieh-En Tsai and Jean Oh. 2020. A Generative Approach for Socially Compliant Navigation. In *2020 IEEE International Conference on Robotics and Automation (ICRA)* (Paris, France). IEEE Computer Society, Los Alamitos, CA, USA, 2160–2166. https://doi.org/10.1109/ICRA40945.2020.9197497

[5] Pei Xu, Jean-Bernard Hayet, and Ioannis Karamouzas. 2022. Socialvae: Human trajectory prediction using timewise latents. In *Computer Vision–ECCV 2022, Proceedings* (Tel Aviv, Israel). Springer, 511–528. https://doi.org/10.1007/978-3-031-19772-7_30

[6] Zikang Zhou, Luyao Ye, Jianping Wang, Kui Wu, and Kejie Lu. 2022. Hivt: Hierarchical vector transformer for multi-agent motion prediction. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition* (New Orleans, USA). IEEE Computer Society, Los Alamitos, CA, USA, 8823–8833. https://doi.org/10.1109/CVPR52688.2022.00862

Received 20 February 2007; revised 12 March 2009; accepted 5 June 2009