# Rethinking of learning-based Pedestrian Trajectory Prediction for social-compliant Motion Planning in a human-centred Environment

Haoming Zhang
Christoph Henke
Frank Hees
{haoming.zhang,christoph.henke,frank.hees}@ifu.rwth-aachen.de
Institute for Management Cybernetics
Aachen, Germany

## ABSTRACT

In the domain of Autonomous Navigation of Robots, Trajectory prediction of the surrounding pedestrians becomes essential for reliable Path and Motion planning in a crowded human-centred scenario. Predicting the future trajectory of pedestrian remains a challenging problem due to unknown inherent dynamics, social behaviors and uncertainties in the environment. Of all the different modelling strategies, learning-based algorithms have shown considerable capability as they naturally capture the human behaviors in a crowded scenario. By exploring the past trajectories and optionally the cues from the environment, the future positions of pedestrian trajectory on a pre-defined horizon could be predicted sequentially. In this research, we focus on the question, **How to model the human interactions more efficiently to benefit the motion planning of mobile robots?** We propose our ideas based on previous work by exploiting Graph theory and Recurrent Neural Networks (RNN).

## CCS CONCEPTS

• **Computing methodologies → Machine learning approaches**.

## KEYWORDS

learning-based, pedestrian, social-aware, trajectory prediction, motion planning

## 1 INTRODUCTION

Understanding the social behavior of humans has kept drawing attention in the last few decades while developing of autonomous systems [5, 13, 16, 22]. In several applications such as autonomous driving [20], mobile robots [31], surveillance [9], the state-of-the-art approaches have shown high potential to improve the social compliance for breaking through the dead end (local minimum) of the system functionalities in a crowed urban environment.

As an example of mobile robot in such space, the robot is frequently suffering from the **Freezing Robot Problem (FRP)** [33]. Even novel motion planning algorithms have been exploring new methodologies to control the robot under intrinsic and extrinsic uncertainties and generate hypotheses of feasible path by optimizing goal-oriented target function with contextual constraints [21, 25]. The planed motion could still lose the validity in a human-centred space. In such situations, the robot with an invalid motion plan might not only collide with pedestrians, but also risk the localization, where the pose of the robot cannot be referenced accurately (using GPS, motion capture system) [7]. As mentioned in [33], the key to navigate in a scenario filled with pedestrians could be traced back to the **joint collision avoidance** based on social forces [13]. Observing the same conclusion in many previous work, the awareness of human behavior becomes necessary to plan a more robust motion. With this conclusion, forecasting the human trajectories in a short future is essential to inform the planning algorithm for a smoother robot motion.

However, pedestrian trajectory prediction for robots is a challenging problem inherently. It requires understanding of human motion, intention and interaction with limited sensory inputs [8]. Plenty of approaches have been proposed in recent research, which can be classified in three categories **physics-based, pattern-based** and **planning-based** trajectory prediction according to [26]. In physics-based methods, the motion of dynamic objects and the human social behaviors are often modeled with hand-crafted parametric functions explicitly [6, 13, 28, 34]. By inferring these predefined functions on system state such as velocity or formulating these social factors into a energy-minimizing mechanism, future trajectory of objects could be forecasted upon the physical model based on Newton's laws. Although this kind of method has been exploited in a long time, the performance is still restricted because of limited representation capacity of non-linearity and strong assumptions such as constant velocity [29], linear dynamics [24, 38], no contextual information [1, 11] etc. Considering these limitations, pattern-based (learning-based) approaches abandon the classical pre-defined models and allow learning to implicitly represent dynamics and social features using common approximations such as Gaussian Process (GP) [15], Gaussian Mixture Model (GMM) [17] and neural networks (NN) [10]. In contrast to learning-based prediction, which fits the target function with given data passively, planning-based methods are reasoning the rationality of the models actively by optimizing the action cost or imitating to learn the non-linear target function.

Even though competitive results are shown within these methods, this problem can only be considered solved in fine-tuned, deterministic industrial use-cases, which can hardly be established in an extrinsic scenario due to the uncertainties of the environment [26]. Furthermore, to our knowledge, motion planning algorithms by now are still less benefited from social awareness from a practical point of view, since the previous work are mostly validated in simulation [3].

In this short report, regarding our research question for modelling the social behaviors to predict the pedestrian trajectory, we aim to briefly review learning-based methods using RNNs for pedestrian trajectory prediction and rethink the modelling strategy of social interactions with contextual information and its practicality with a well-defined problem configuration of mobile robots.

This paper is structured as follows: section 2 summarizes the learning based approaches in brief. The problem definition and our research ideas will be introduced in section 3 while conclusions and future plan are presented in section 4.

## 2 LEARNING-BASED TRAJECTORY PREDICTION

Given observations of the sequential pedestrian trajectories and environment representation, learning-based methods propose to extract the motion patterns and joint intention of humans. These methods are furthermore distinguished into **sequential and non-sequential**, **location dependent and independent**, **context dependent and independent** models. Instead of taking the causality between each observed pose on the trajectory into account, common motion patterns will be approximated and clustered to predict the future trajectory within non-sequential modeling [37]. Compared to sequential modeling, less data is required and the learned motion patterns highly depend on the underlying scenario, which typically leads to insufficient generalization. Therefore, sequential modeling with RNNs turns to be more popular in recent research. Alahi et. al 2016 [1] proposed a Long Short Term Memory (LSTM) model to learn the trajectory of each human in continuous space. The social interaction of pedestrians is captured using a social pooling neural layer, which connects the LSTMs so that the hidden states in the LSTMs are shared. Later, this method has been extended to enable the social pooling to capture more social factors like grouping and context awareness [2, 11, 23, 30, 35]. More than pooling layer, spatio-temporal information beside spatial information was also maintained [12, 14, 36] in order to model the temporal attention of human in social interaction.

Along with validation of aforementioned approaches, which are mostly supposed to learn social behaviors without consideration of contextual information, it has been proved in several work that the contextual information is crucial to generalize the trajectory prediction [19, 27, 36]. In [27], the static environmental features are extracted by a common convolution neural network (CNN) and then fed into the LSTM of each human as physical attention. Furthermore, besides social pooling, [19] introduces navigation pooling to indicate the probability of positions while navigating and semantic pooling, which detects the cross-able areas in the current scene. These methods have shown a higher accuracy compared to vanilla LSTMs without contextual information.

## 3 OUR RESEARCH IDEAS

Based on the state-of-the-art methods in the research community, we address this problem in the case of mobile robots and state our research ideas in this section.

### 3.1 Problem Setup and Notation

In our use-case, we setup the problem with following assumptions:

(1) the pedestrians are tracked and re-identified in the crowd and their trajectories (sequential 2D position) in euclidean state space is given in global coordinate system $\mathcal{K}_w$.
(2) the robot is moving and its pose is given in the same coordinate system $\mathcal{K}_w$.

Let $p_t^i = (x_t^i, y_t^i)$ indicate the position of an agent $h^i$, the 2-D sequence $\mathcal{T}_{obs}^i = \{p_1^i, p_2^i, ..., p_{obs}^i\}$ be the trajectory till the timestamp of observation $t_{obs}$.

Following [19], given a semantic map $\mathcal{I}_{sem} \in \mathbb{R}^{n \times n}$, which labels the object boundaries projected on the ground in euclidean space, and a navigation map $\mathcal{I}_{nav} \in \mathbb{R}^{n \times n}$, which analyses the probability of the positions to be visited and trajectories of neighboring agents $\mathcal{N}_{traj}^i = \{\mathcal{T}_{obs}^1, \mathcal{T}_{obs}^2, ..., \mathcal{T}_{obs}^{i-1}\}$, the features used to model the social interaction should be extracted as

$$\mathcal{F}^i = \mathcal{G}(\mathcal{I}_{sem}, \mathcal{I}_{nav}, \mathcal{N}_{traj}^i)$$

and later the interaction for the human agent $h^i$ is supposed to be modeled as

$$\mathcal{S}_i = RNN(\mathcal{F}^i, \mathcal{T}_i)$$

Finally, with a trajectory approximator such as bivariate Gaussian distribution [36], the future trajectory could be approximated as

$$\mathcal{T}_{obs+t}^i = \{p_{obs+1}^i, p_{obs+2}^i, ..., p_{obs+t}^i\} = \mathcal{P}(\mathcal{S}_i)$$

### 3.2 Our Ideas

Rethinking of learning-based methods by hybrid-modelling with RNNs and graph theory shows a great potential to predict human future trajectory resulting in better motion planning of mobile robots.

Inspired from [19] and [36], our first ideas and interests include the following options:

(1) Factor graph, a derivative of Bayesian network is often used to model the target function in state estimation domain for the reason that it enables us to solve the optimization problem with constraints [4, 32]. In a factor graph, the nodes usually present the optimization target and the factors describe the constraints between two nodes. With a factor graph, social relations could be modeled as a dynamic graph, see figure 1, (a). Based on [19], more inherent properties of factor graph could be used while modelling the social interaction, such as elimination (figure 1, (b)), which clusters and reforms the factors between nodes to keep the graph sparse.
(2) New research idea relaid on reasoning the human behaviors to predict the future has been proposed [18]. Using factor graph, novel factors based on the reasoning could be used to enlarge the capability of social modelling.
(3) Although the effectiveness of the pooling mechanism in capturing the social interaction in sequential modelling has been validated in previous work, modelling strategy of social behavior still needs to be discussed in the presence of contextual information. As shown in figure 1, (c), apart from the positions on trajectories having different state space, also the contextual information could have different state

spaces. A semantic map usually represents the geometric and characteristic information of environment objects, while the navigation map models the probabilities of locations to be visited. As a result of larger state space, Formulating the translation and causality from the trajectory and contextual information makes it a problem in higher-dimensional space. Normally, instead of learning the causality with all of these information, attention-based mechanisms are used to weigh different features in an implicit way. In our research, we also rely on social attention, but different attention forms will be modeled before a normalized fusion, which was proposed in [36].

(4) Many concerns have to be addressed in integrating pedestrian trajectory prediction in motion planning of mobile robots, especially the practical implementation on mobile robots, whose computation resource is often limited. The trade-off of prediction quality and efficiency will be discussed in our research work. Within the application of modelling social interaction for mobile robots, a socially-aware path and motion planner will be proposed.

## 4 CONCLUSION

In this report, we reviewed the state-of-the-art approaches for pedestrian trajectory prediction. We also stated the necessity of modelling the social interaction for robust motion and path planning. With this conclusion, we presented our research interest and ideas to solve this problem in the use-case of mobile delivery robot. In the future, the proposed ideas will be further implemented and evaluated. We also plan to optimize the current motion planning algorithm with pedestrian trajectory prediction as constraints and study their suitability on real robot system.

## ACKNOWLEDGMENTS

This work has been fully funded by the German **Federal Ministry of Education and Research (BMBF)** within the project **16SV8232** 'Urbane, automatisierte, nutzerorientierte Transportplatform (Phase II)'.

## REFERENCES

[1] Alexandre Alahi, Kratarth Goel, Vignesh Ramanathan, Alexandre Robicquet, Fei-Fei Li, and Silvio Savarese. 2016. Social LSTM: Human Trajectory Prediction in Crowded Spaces. *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (2016), 961–971.

[2] Federico Bartoli, Giuseppe Lisanti, Lamberto Ballan, and Alberto Del Bimbo. 2017. Context-Aware Trajectory Prediction. *2018 24th International Conference on Pattern Recognition (ICPR)* (2017), 1941–1946.

[3] Yu Fan Chen, Miao Liu, Michael Everett, and Jonathan P. How. 2016. Decentralized non-communicating multiagent collision avoidance with deep reinforcement learning. *2017 IEEE International Conference on Robotics and Automation (ICRA)* (2016), 285–292.

[4] Frank Dellaert and Michael Kaess. 2017. Factor Graphs for Robot Perception. *Foundations and Trends in Robotics* 6 (01 2017), 1–139. https://doi.org/10.1561/2300000043

[5] A. Elnagar and K. Gupta. 1998. Motion prediction of moving objects based on autoregressive model. *IEEE Transactions on Systems, Man, and Cybernetics - Part A: Systems and Humans* 28, 6 (Nov 1998), 803–810. https://doi.org/10.1109/3468.725351

[6] Ashraf Elnagar and K. Gupta. 1998. Motion prediction of moving objects based on autoregressive model. *IEEE Trans. Systems, Man, and Cybernetics, Part A* 28 (1998), 803–810.

[7] T. Fan, X. Cheng, J. Pan, P. Long, W. Liu, R. Yang, and D. Manocha. 2019. Getting Robots Unfrozen and Unlost in Dense Pedestrian Crowds. *IEEE Robotics and Automation Letters* 4, 2 (April 2019), 1178–1185. https://doi.org/10.1109/LRA.2019.2891491

[8] Sarah Ferguson, Brandon Luders, Robert Grande, and Jonathan How. 2014. Real-Time Predictive Modeling and Robust Avoidance of Pedestrians with Uncertain, Changing Intentions. (04 2014). https://doi.org/10.1007/978-3-319-16595-0_10

[9] Tharindu Fernando, Simon Denman, Sridha Sridharan, and Clinton Fookes. 2019. Pedestrian Trajectory Prediction with Structured Memory Hierarchies. In *Machine Learning and Knowledge Discovery in Databases*, Michele Berlingerio, Francesco Bonchi, Thomas Gärtner, Neil Hurley, and Georgiana Ifrim (Eds.). Springer International Publishing, Cham, 241–256.

[10] Michael Goldhammer, Konrad Doll, Ulrich Brunsmann, Andre Gensler, and Bernhard Sick. 2014. Pedestrian's Trajectory Forecast in Public Traffic with Artificial Neural Networks. *2014 22nd International Conference on Pattern Recognition* (2014), 4110–4115.

[11] Agrim Gupta, Justin Johnson, Li Fei-Fei, Silvio Savarese, and Alexandre Alahi. 2018. Social GAN: Socially Acceptable Trajectories with Generative Adversarial Networks. *CoRR* abs/1803.10892 (2018). arXiv:1803.10892 http://arxiv.org/abs/1803.10892

[12] Sirin Haddad, Meiqing Wu, Wei He, and Siew Kei Lam. 2019. Situation-Aware Pedestrian Trajectory Prediction with Spatio-Temporal Attention Model. https://doi.org/10.3217/978-3-85125-652-9

[13] Dirk Helbing and Peter Molnar. 1998. Social Force Model for Pedestrian Dynamics. *Physical Review E* 51 (05 1998). https://doi.org/10.1103/PhysRevE.51.4282

[14] Yingfan Huang, Huikun Bi, Zhaoxin Li, Tianlu Mao, and Zhaoqi Wang. [n.d.]. STGAT : Modeling Spatial-Temporal Interactions for Human Trajectory Prediction.

[15] Joshua Joseph, Finale Doshi-Velez, Albert S. Huang, and Nicholas Roy. 2011. A Bayesian nonparametric approach to modeling motion patterns. *Autonomous Robots* 31, 4 (19 Aug 2011), 383. https://doi.org/10.1007/s10514-011-9248-x

[16] Eckhard Kruse and Friedrich M. Wahl. 1998. Camera-based observation of obstacle motions to derive statistical data for mobile robot motion planning. *Proceedings. 1998 IEEE International Conference on Robotics and Automation (Cat. No.98CH36146)* 1 (1998), 662–667 vol.1.

[17] Tomasz Piotr Kucner, Martin Magnusson, Erik Schaffernicht, Victor Manuel Hernandez Bennetts, and Achim J. Lilienthal. 2017. Enabling Flow Awareness for Mobile Robots in Partially Observable Environments. *IEEE Robotics and Automation Letters* 2 (2017), 1093–1100.

[18] J. Liang, L. Jiang, J. C. Niebles, A. G. Hauptmann, and L. Fei-Fei. 2019. Peeking Into the Future: Predicting Future Person Activities and Locations in Videos. In *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. 5718–5727. https://doi.org/10.1109/CVPR.2019.00587

[19] Matteo Lisotto, Pasquale Coscia, and Lamberto Ballan. 2019. Social and Scene-Aware Trajectory Prediction in Crowded Spaces. *ArXiv* abs/1909.08840 (2019).

[20] Y. Luo, P. Cai, A. Bera, D. Hsu, W. S. Lee, and D. Manocha. 2018. PORCA: Modeling and Planning for Autonomous Driving Among Many Pedestrians. *IEEE Robotics and Automation Letters* 3, 4 (Oct 2018), 3418–3425. https://doi.org/10.1109/LRA.2018.2852793

[21] L. Palmieri, T. P. Kucner, M. Magnusson, A. J. Lilienthal, and K. O. Arras. 2017. Kinodynamic motion planning on Gaussian mixture fields. In *2017 IEEE International Conference on Robotics and Automation (ICRA)*. 6176–6181. https://doi.org/10.1109/ICRA.2017.7989731

[22] Alex Pentland and Andrew Liu. 1997. Modeling and Prediction of Human Behavior. *Neural Computation* 11 (09 1997). https://doi.org/10.1162/089976699300016890

[23] Mark Pfeiffer, Giuseppe Paolo, Hannes Sommer, Juan I. Nieto, Roland Siegwart, and Cesar Cadena. 2017. A Data-driven Model for Interaction-Aware Pedestrian Motion Prediction in Object Cluttered Environments. *2018 IEEE International Conference on Robotics and Automation (ICRA)* (2017), 1–8.

[24] Ewoud A. I. Pool, Julian F. P. Kooij, and Dariu Gavrila. 2017. Using road topology to improve cyclist path prediction. *2017 IEEE Intelligent Vehicles Symposium (IV)* (2017), 289–296.

[25] Christoph Rösmann, Frank Hoffmann, and Torsten Bertram. 2017. Integrated online trajectory planning and optimization in distinctive topologies. *Robotics and Autonomous Systems* 88 (2017), 142–153.

[26] Andrey Rudenko, Luigi Palmieri, Serge Herman, Kris M. Kitani, Dariu M. Gavrila, and Kai Oliver Arras. 2019. Human Motion Trajectory Prediction: A Survey. *ArXiv* abs/1905.06113 (2019).

[27] Amir Sadeghian, Vineet Kosaraju, Ali Sadeghian, Noriaki Hirose, Hamid Rezatofighi, and Silvio Savarese. 2019. SoPhie: An Attentive GAN for Predicting Paths Compliant to Social and Physical Constraints. 1349–1358. https://doi.org/10.1109/CVPR.2019.00144

[28] Nicolas Schneider and Dariu M. Gavrila. 2013. Pedestrian Path Prediction with Recursive Bayesian Filters: A Comparative Study. In *Pattern Recognition*, Joachim Weickert, Matthias Hein, and Bernt Schiele (Eds.). Springer Berlin Heidelberg, Berlin, Heidelberg, 174–183.

[29] Christoph Schöller, Vincent Aravantinos, Florian Lay, and Alois Knoll. 2019. The Simpler the Better: Constant Velocity for Pedestrian Motion Prediction. *ArXiv* abs/1903.07933 (2019).

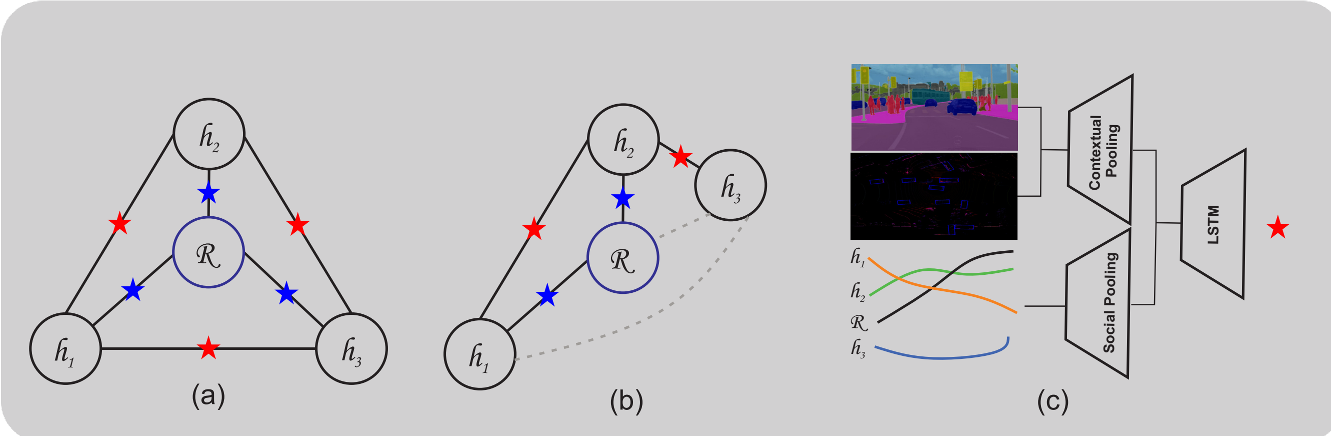


**Figure 1: a: proposed factor graph, red star for factor of human-human interaction, blue star for factor of human-robot interaction, b: factor graph after elimination, c: considered feature extraction to model the social interaction, above: semantic grid, middle: navigation grid, below: trajectories**


[30] Kevin Sun, Zhi Yan, Sergi Mellado, Marc Hanheide, and Tom Duckett. 2018. 3DOF Pedestrian Trajectory Prediction Learned from Long-Term Autonomous Mobile Robot Deployment Data. https://doi.org/10.1109/ICRA.2018.8461228

[31] L. Sun, Z. Yan, S. M. Mellado, M. Hanheide, and T. Duckett. 2018. 3DOF Pedestrian Trajectory Prediction Learned from Long-Term Autonomous Mobile Robot Deployment Data. In *2018 IEEE International Conference on Robotics and Automation (ICRA)*. 5942–5948. https://doi.org/10.1109/ICRA.2018.8461228

[32] Duy-Nguyen Ta, Marin Kobilarov, and Frank Dellaert. 2014. A factor graph approach to estimation and model predictive control on Unmanned Aerial Vehicles. *2014 International Conference on Unmanned Aircraft Systems (ICUAS)* (2014), 181–188.

[33] Pete Trautman and Andreas Krause. 2010. Unfreezing the Robot: Navigation in Dense, Interacting Crowds. *Proceedings of the ... IEEE/RSJ International Conference on Intelligent Robots and Systems. IEEE/RSJ International Conference on Intelligent Robots and Systems*, 797–803. https://doi.org/10.1109/IROS.2010.5654369

[34] Jur van den Berg, Ming Lin, and Dinesh Manocha. 2008. Reciprocal Velocity Obstacles for Real-Time Multi-agent Navigation. *ICRA*, 1928–1935. https://doi.org/10.1109/ROBOT.2008.4543489

[35] Daksh Varshneya and G. Srinivasaraghavan. 2017. Human Trajectory Prediction using Spatially aware Deep Attention Models. *ArXiv* abs/1705.09436 (2017).

[36] Anirudh Vemula, Katharina Muelling, and Jean Oh. 2017. Social Attention: Modeling Attention in Human Crowds. *2018 IEEE International Conference on Robotics and Automation (ICRA)* (2017), 1–7.

[37] Youngjoon Yoo, Kimin Yun, Sangdoo Yun, Jonghee Hong, Hawook Jeong, and Jin Choi. 2016. Visual Path Prediction in Complex Scenes with Crowded Moving Objects. 2668–2677. https://doi.org/10.1109/CVPR.2016.292

[38] Bolei Zhou, Xiaoou Tang, and Xiaogang Wang. 2015. Learning Collective Crowd Behaviors with Dynamic Pedestrian-Agents. *International Journal of Computer Vision* 111, 1 (01 Jan 2015), 50–68. https://doi.org/10.1007/s11263-014-0735-3