Intent-Aware Pedestrian Prediction for Adaptive Crowd Navigation

Kapil D. Katyal, I-Jeng Wang, Gregory D. Hager, and Chien-Ming Huang

ABSTRACT

In this article, we describe the work of a team of researchers from the Johns Hopkins University Applied Physics Laboratory (APL) and Johns Hopkins University (JHU) to develop adaptive crowd navigation policies for robots by reasoning and predicting future pedestrian motion.

INTRODUCTION

Humans and other biological agents have innate abilities to consistently predict the effects of their own actions as well as those of actions of other agents in the environment. This innate ability to predict allows for robust planning in highly dynamic and uncertain situations. We contrast this with existing intelligent artificial agents that have a limited ability to anticipate and predict beyond what the sensor can see. Our belief is that this prediction capability is critical to advance next-generation intelligent systems for safe and robust operations in dynamic environments with (uncoordinated) human activities.

In this article, we describe an approach that provides mobile robotic systems the ability to predict and anticipate motion of pedestrians and other dynamic obstacles in the environment. Specifically, we investigate human motion in crowded spaces to explore how to recognize pedestrians' navigation intent, how to predict pedestrians' motion, and how a robot may adapt its navigation policy dynamically when facing unexpected human movements. We experimentally demonstrate the effectiveness of our prediction algorithm using real-world data sets on pedestrians and achieve comparable or better prediction accuracy compared with several state-of-the-art approaches (shown in Table 1). Moreover, we show that confidence in the prediction of pedestrian motion can be used to adjust the risk of a navigation policy adaptively to afford the most comfortable level as measured by the frequency of personal space violation in comparison with baselines. Furthermore, our adaptive navigation policy is able to reduce the number of collisions by 43% in the presence of novel pedestrian motion not seen during training.

TECHNICAL APPROACH

Machine learning has had a significant impact on many domains, including object recognition, natural language processing, and speech recognition. In recent years, we have seen a significant rise in the use of machine learning, and reinforcement learning specifically, for robotic navigation tasks. The advantage is that robotic systems are now capable of *learning* skills versus

This article draws heavily, reproducing some content verbatim, from Ref. 1: K. Katyal, G. D. Hager, and C.-M. Huang, "Intent-aware pedestrian prediction for adaptive crowd navigation," in *Proc. 2020 Int. Conf. Robot. Automat. (ICRA)*, pp. 3277–3283, https://doi.org/10.1109/ICRA40945.2020.9197434. Additional technical information can be found in the ICRA article.

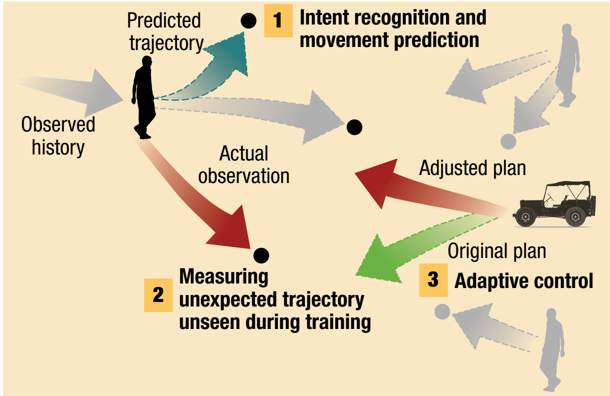


Figure 1. Our approach for adaptive crowd navigation. (1) We first estimate the intent of the pedestrian and combine the estimation with past trajectories to predict future motion. (2) We then measure the error between our prediction and reality to develop an adaptive mobile robot controller. (3) The adaptive mobile robot controller switches between aggressive and risk-averse policies based on accuracy of the prediction. (© 2020 IEEE. Modified and reprinted, with permission, from Ref. 1.)

being preprogrammed to perform tasks. This results in systems that are more adaptive and capable of being deployed in unstructured environments. Despite these advancements, many challenges remain with machine learning–based solutions. In particular, algorithms are typically trained in a certain environment or a specific data set. When faced with situations that are significantly outside of the distribution they were originally trained on, these approaches struggle.

In the work described in this article, we develop an adaptive crowd navigation policy that is robust to changes in the distribution from the originally trained policy. Our approach for adaptive crowd navigation, as described in Figure 1, focuses on three stages. We first estimate a probabilistic representation of navigation intent and use this estimated intent to predict future motion. The second step is to observe measurement errors between the estimated motion and observed motion as a heuristic for out-of-distribution events. Our hypothesis is that prediction can play a large role in determining whether the robot is encountering a novel situation. If the predicted motion is highly correlated with the observed motion, we use an aggressive policy that allows the robot to reach the destination as quickly as possible. If the motion is uncorrelated, we revert to a risk-averse controller that favors conservative motion to avoid collisions.

Pedestrian Prediction

In this section, we describe our approach to estimating intent and predicting future motion. As shown in Figure 2, this consists of a neural network architecture that combines observed trajectories with a probabilistic representation of intent to estimate future motion.

We compare our pedestrian prediction algorithm with several baselines and show that the average displacement error (ADE) and final displacement error (FDE) of our predicted trajectory meets or exceeds prior state-of-the-art results (Table 1).

Adaptive Crowd Navigation

Reinforcement learning has made significant strides in allowing robotic systems to learn capabilities versus requiring preprogrammed routines. One of the challenges to existing state-of-the-art reinforcement learning–based navigation policies is robustness to changing distributions in observations from the training data. To illustrate this, we conducted a series of experiments where we changed the distribution of pedestrian motion during testing of the reinforcement learning policy. The

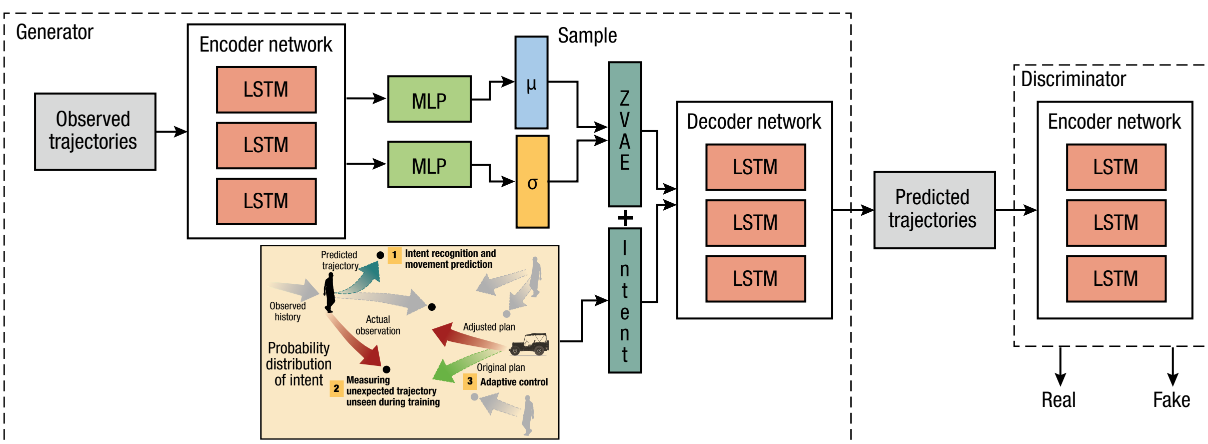


Figure 2. Pedestrian prediction neural network architecture. We first estimate a probability distribution of the intent of the pedestrian and then combine that estimation with past trajectories using long short-term memory (LSTM) and multilayer perceptron (MLP) neural networks. The network learns a Gaussian representation that can be sampled from and combined with the intent to predict future pedestrian motion. (© 2020 IEEE. Modified and reprinted, with permission, from Ref. 1.)

Table 1. Results of our pedestrian prediction approach using real-world data sets and showing reduced error in our prediction compared to several state-of-the-art pedestrian prediction algorithms

Metric	Data Set[a]	Linear	LSTM	S-LSTM[b]	SoPhie 1V-20[c]	SGAN[d] 1V-1	SGAN[d] 1V-20	SGAN[d] 20V-20	SGAN[d] 20VP-20	Ours 1V-1	Ours 1V+ IR-1	Ours 1V+ IR-20	Ours 20V+ IR-20
ADE	ETH	1.33	1.09	1.09	0.70	1.13	1.03	0.81	0.87	0.96	0.85	0.77	**0.69**
	Hotel	**0.39**	0.86	0.79	0.76	1.01	0.90	0.72	0.67	0.60	0.48	0.42	**0.39**
	Univ	0.82	0.61	0.67	0.54	0.60	0.58	0.60	0.76	0.55	0.53	**0.51**	0.56
	Zara1	0.62	0.41	0.47	**0.30**	0.42	0.38	0.34	0.35	0.45	0.41	0.36	0.35
	Zara2	0.77	0.52	0.56	0.38	0.52	0.47	0.42	0.42	0.38	0.33	**0.30**	0.31
	Average	0.79	0.70	0.72	0.54	0.74	0.67	0.58	0.61	0.59	0.52	0.47	**0.46**
FDE	ETH	2.94	2.41	2.35	1.43	2.21	2.02	1.52	1.62	1.85	1.80	1.66	**1.42**
	Hotel	**0.72**	1.91	1.76	1.67	2.18	1.97	1.61	1.37	1.18	1.04	0.94	0.79
	Univ	1.59	1.31	1.40	1.24	1.28	1.22	1.26	1.52	1.17	1.13	**1.09**	1.17
	Zara1	1.21	0.88	1.00	**0.63**	0.91	0.84	0.69	0.68	0.94	0.87	0.79	0.74
	Zara2	1.48	1.11	1.17	0.78	1.11	1.01	0.84	0.84	0.79	0.72	**0.65**	0.66
	Average	1.59	1.52	1.54	1.15	1.54	1.41	1.18	1.21	1.19	1.11	1.03	**0.96**

Smallest error rates are in bold.

[a] Publicly available ETH (see Ref. 2) and UCY (see Ref. 3) repositories. [b] See Ref. 4. [c] See Ref. 5. [d] See Ref. 6.

result shows a significant increase in the number of collisions when presented with novel pedestrian motion. Our adaptive crowd navigation policy uses prediction as a criterion for detecting novel pedestrian motion. The underlying intuition behind our approach is similar to how we believe humans approach navigation in complex scenes. If our prediction matches reality, we maintain an aggressive navigation policy. Conversely, if our prediction is no longer accurate, we develop a risk-averse approach and navigate cautiously through the environment. By developing an adaptive, risk-sensitive controller based on predicting pedestrian motion, we are able to demonstrate a significant reduction in the number of collisions compared to state-of-the-art navigation policies, as shown in Table 2.

CONCLUSION

In this article, we describe a novel approach to estimating pedestrian intent and use pedestrian intent to make better predictions of pedestrian motion. Further, we show that errors in pedestrian motion can be used

Table 2. Results of our adaptive mobile robot controller in the presence of novel pedestrian motion

Method	Distribution Shift	Successful Trials	No. Collisions	Time-outs	Navigation Time	Discomfort Rate	Average Reward
CADRL-5[a]	N	455	45	0	4.48	2.02	0.349
SARL-5[b]	N	490	5	5	4.61	0.99	0.389
CADRL-5	Y	420	80	**0**	**4.52**	3.53	0.296
SARL-5	Y	425	70	5	4.62	2.27	0.303
SVM-A-5[c]	Y	426	68	6	5.34	2.14	0.331
SGAN-A-5[d]	Y	445	45	10	6.31	2.03	0.386
Ours-A-5	Y	**450**	**40**	10	6.74	**1.98**	**0.409**
SARL-10	Y	388	99	13	**5.21**	2.62	0.234
Ours-A-10	Y	**444**	**54**	**2**	8.49	**2.18**	**0.330**
SARL-15	Y	290	205	5	**5.30**	4.89	0.115
Ours-A-15	Y	**366**	**132**	**2**	8.69	**4.20**	**0.212**
SARL-20	Y	172	324	4	**5.27**	6.70	–0.017
Ours-A-20	Y	**262**	**237**	**1**	8.65	**6.29**	**0.066**

Boldface denotes best results.

[a] See Ref. 7. [b] See Ref. 8. [c] See Ref. 9. [d] See Ref. 4.

to alter the risk of a mobile robot navigating in the presence of pedestrians. Finally, we show that a risk-sensitive, adaptive motion planner can significantly reduce the number of collisions, particularly in the presence of novel pedestrian motion. Our future efforts include modeling complex social interactions, behaviors, and personalities to improve socially aware navigation and developing continually learning policies that improve navigation strategies over time.

REFERENCES

[1]K. Katyal, G. D. Hager, and C.-M. Huang, "Intent-aware pedestrian prediction for adaptive crowd navigation," in *Proc. 2020 Int. Conf. Robot. Automat. (ICRA)*, pp. 3277–3283, https://doi.org/10.1109/ICRA40945.2020.9197434.

[2]S. Pellegrini, A. Ess, and L. Van Gool, "Improving data association by joint modeling of pedestrian trajectories and groupings," in *Computer Vision – ECCV 2010*, K. Daniilidis, P. Maragos, and N. Paragios, Eds. Berlin, Heidelberg: Springer Berlin Heidelberg, 2010, pp. 452–465.

[3]A. Lerner, Y. Chrysanthou, and D. Lischinski, "Crowds by example," *Comput. Graph. Forum*, vol. 26, no. 3, pp. 655–664, 2007. https://doi.org/10.1111/j.1467-8659.2007.01089.x.

[4]A. Alahi, K. Goel, V. Ramanathan, A. Robicquet, L. Fei-Fei, and S. Savarese, "Social LSTM: Human trajectory prediction in crowded spaces," in *Proc. 2016 IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, 2016, pp. 961–971, https://doi.org/10.1109/CVPR.2016.110.

[5]A. Sadeghian, V. Kosaraju, A. Sadeghian, N. Hirose, H. Rezatofighi, and S. Savarese, "SoPhie: An attentive GAN for predicting paths compliant to social and physical constraints," in *Proc. 2019 IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, 2019, pp. 1349–1358, https://doi.org/10.1109/CVPR.2019.00144.

[6]A. Gupta, J. Johnson, L. Fei-Fei, S. Savarese, and A. Alahi, "Social GAN: Socially acceptable trajectories with generative adversarial networks," in *Proc. 2018 IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, 2018, pp. 2255–2264, https://doi.org/10.1109/CVPR.2018.00240.

[7]Y. F. Chen, M. Liu, M. Everett, and J. How, "Decentralized non-communicating multiagent collision avoidance with deep reinforcement learning," in *Proc. 2017 Int. Conf. Robot. Automat. (ICRA)*, pp. 285–292, https://doi.org/10.1109/ICRA.2017.7989037.

[8]C. Chen, Y. Liu, S. Kreiss, and A. Alahi, "Crowd-robot interaction: Crowd-aware robot navigation with attention-based deep reinforcement learning," in *Proc. 2019 Int. Conf. Robot. Automat. (ICRA)*, pp. 6015–6022, doi: 10.1109/ICRA.2019.8794134..

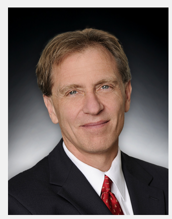
[9]R.-E. Fan, K.-W. Chang, C.-J. Hsieh, X.-R. Wang, and C.-J. Lin, "Liblinear: A library for large linear classification," *J. Mach. Learn. Res.*, vol. 9, pp. 1871–1874, Jun. 2008, http://dl.acm.org/citation.cfm?id=1390681.1442794.

Kapil D. Katyal, Research and Exploratory Development Department, Johns Hopkins University Applied Physics Laboratory, Laurel, MD

Kapil D. Katyal is a project manager and supervisor of the Robotic Intelligence Section in APL's Research and Exploratory Development Department. He has a BS in computer science from Pennsylvania State University (Penn State) and an MS in electrical engineering from Columbia University. Kapil has 13 years of artificial intelligence (AI) and robotics experience and has led projects ranging from upper limb prosthetics, neural interfaces, retinal prosthetics, and robot navigation. His research interests span the areas of machine learning with robotic systems. His email address is kapil.katyal@jhuapl.edu.

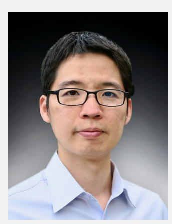
I-Jeng Wang, Research and Exploratory Development Department, Johns Hopkins University Applied Physics Laboratory, Laurel, MD

I-Jeng Wang is a project manager and chief scientist in APL's Research and Exploratory Development Department. He has a BS in engineering science from National Chiao Tung University, an MS in electrical engineering from Pennsylvania State University (Penn State), and a PhD in electrical engineering from Purdue University. He has a broad background in areas including stochastic control and optimization, Bayesian networks, resource allocation, and machine learning. He has led multiple research and development projects funded by the Defense Advanced Research Projects Agency (DARPA), the Office of Naval Research (ONR), the Army Research Laboratory (ARL), and the National Science Foundation (NSF), conducting research in compressed sensing, machine learning algorithms, and Bayesian inference and decision theory. His email address is i-jeng.wang@jhuapl.edu.



Gregory D. Hager, Johns Hopkins University, Baltimore, MD

Gregory D. Hager is the Mandell Bellmore Professor of Computer Science at Johns Hopkins University. He holds a BA in mathematics from Luther College and an MSE and a PhD in computer science from the University of Pennsylvania. His principal areas of research are computer vision and robotics, with application interests in the area of medical devices and human–machine systems for manufacturing or in-home applications. Gregory directs the Computational Interaction and Robotics Lab (CIRL) in the Laboratory for Computational Sensing and Robotics and is chair of the Computing Community Consortium Council, a member of the boards of the Computing Research Association and the International Federation of Robotics Research, and a Fellow of the IEEE. His email address is hager@cs.jhu.edu.



Chien-Ming Huang, Johns Hopkins University, Baltimore, MD

Chien-Ming Huang is a John C. Malone Assistant Professor in the Department of Computer Science at Johns Hopkins University (JHU), where he studies human–machine teaming and creates innovative, intuitive, personalized technologies to provide social, physical, and behavioral support for people with a variety of abilities and characteristics, including children with autism spectrum disorders. Huang directs JHU's interdisciplinary Intuitive Computing Laboratory and is a member of JHU's Malone Center for Engineering in Healthcare and the Laboratory for Computational Sensing and Robotics. He has won several publication awards. His email address is cmhuang@cs.jhu.edu.