Improving the Robustness of Social Robot Navigation Systems

Nathan Tsoi
Yale University
nathan.tsoi@yale.edu

Marynel Vázquez
Yale University
New Haven, CT, USA

ABSTRACT

Our aim is to advance the reliability of autonomous social navigation. We have researched how simulation may advance this goal via crowdsourcing. We recently proposed the Simulation Environment for Autonomous Navigation (SEAN) and deployed it at scale on the web to quickly collect data via the SEAN Experimental Platform (SEAN-EP). Using this platform, we studied participants' perceptions of a robot when seen in a video versus interacting with it in simulation. Our current research builds on this prior work to make autonomous social navigation more reliable by classifying and automatically detecting navigation errors.

ACM Reference Format:
Nathan Tsoi and Marynel Vázquez. 2021. Improving the Robustness of Social Robot Navigation Systems. In *Companion of the 2021 ACM/IEEE International Conference on Human-Robot Interaction (HRI '21 Companion), March 8–11, 2021, Boulder, CO, USA.* ACM, New York, NY, USA, 3 pages. https://doi.org/10.1145/3434074.3446360

1 INTRODUCTION

Our research in social robot navigation is motivated by concepts from software engineering. Reliability is a crucial component of production-ready software systems. The related measure of availability is often reported in "nines of availability" [12]. For example, 5 nines corresponds to "99.999%" availability which equates to a maximum of 5.26 minutes of downtime per year. Today, users of systems such as social networking platforms and bank websites have an expectation of perpetual availability. More subtle is their expectation that these systems will function without errors. For example, users would not accept an online banking system that even occasionally transferred money to the wrong account due to a system error. The failure of a system to perform its required functions is a failure of the system's reliability [1]. In our view, current social robot navigation methods are far from the high level of availability and reliability demanded by real users. This has inspired our research question: **how can we increase the reliability of social robot navigation methods?** Our key insights are that *(1) we can leverage simulations to gather implicit and explicit human supervision data for autonomous operation; and (2) we can leverage this supervision to autonomously predict unsafe actions for a robot.* We realized these insights in two key ways that advance community efforts. First, we proposed a simulation platform for the fair comparison of social navigation algorithms. Second, we deployed our simulation platform at scale to collect qualitative human feedback. Our future work builds on these advancements. We are currently working to use our simulation to gather human feedback at scale and train autonomous methods for predicting unsafe robot actions.

HRI '21 Companion, March 8–11, 2021, Boulder, CO, USA

ACM ISBN 978-1-4503-8290-8/21/03.
https://doi.org/10.1145/3434074.3446360

Figure 1: SEAN rendering (left to right) a lab scene, warehouse, and outdoor environment. All scenes include dynamic pedestrians and a robot controlled via ROS.

1.1 Benchmarking Social Navigation Methods

There is currently no agreed-upon benchmarking methodology for social robot navigation. One general challenge associated with benchmarks is that they can potentially hinder progress in basic science. This is a problem when the benchmark becomes an end in itself, as opposed to a means of measuring progress. The field of computer vision is exemplary both in this challenge, but also in the progress resulting from benchmarks. Drawing a parallel from these challenges to robotics is complicated by the difference between evaluating performance on static datasets versus working with interactive robotic systems. Simulation can help bridge this gap.

While not a perfect solution, as simulators are not equivalent to the real world, they provide benefits such as accessibility and repeatability. These properties may enable community-wide evaluation of social navigation systems. Our prior work on an evaluation platform for social navigation systems in simulation, SEAN (Fig. 1) [16], is a step towards creating a repeatable, interactive benchmark for social robotic navigation. We hope this will help enable standard evaluation protocols and promote discussion of evaluation methodologies. Awareness of common pitfalls can help the community avoid the potential downsides associated with such benchmarks.

SEAN allows for control of a mobile robot in a virtual and dynamic human environment [16]. It is built on the Unity 3D game engine [5] and interfaces with the Robot Operating System (ROS) [13]. Unity implements the NVIDIA PhysX physics engine, which has been shown to be a promising approach for robot simulation [11]. ROS and Unity communicate via Rosbridge [3], implemented via the ROS# library [2]. The architecture of the system balances between a) ease of integration with existing navigation systems via ROS, b) high visual fidelity capable of creating immersive environments that enable vision-based navigation methods, and c) a cross-platform ecosystem that supports iterative development.

The key design tenets of the simulation platform are usability and flexibility. To this end, we provide a set of scenes, robots, and evaluation metrics within the platform to enable users to use the system with minimal preliminary work. Additionally, we have created an open source repository and supporting documentation to allow the community to improve the simulation environment [4].

Scenes are 3D environments in which a robot operates. SEAN provides four high-fidelity scenes: a replica of the Cocktail party environment from [19], a lab environment, a warehouse, and a larger outdoor city scene. We recently used the first environment to study human perception of spatial robot behavior during situated social interactions [9]. The other three are depicted in (Fig. 1). Because humans play a key part in the study of social robot navigation, each scene includes start and goal positions for human agents to navigate. The simulator uses a hierarchical method of control for the pedestrians which combines Unity's path planning algorithm with a lower-level Social Forces model [15]. The appropriate number of agents can be selected given the size and context of the scene.

Evaluation Toolkit: Central to the goal of benchmarking, SEAN provides the Trial Runner, which enables repeatable and automatic execution of navigation tasks. The Trial Runner performs a *trial* by executing a collection of point-to-point navigation *episodes*. Initial conditions for each episode are random by default, and recorded at the beginning of an episode. This allows episode replay and comparisons of navigation methods under the same initial conditions. The goal of the toolkit is not to enforce any specific metrics for comparison, but enable the community to begin to establish a common evaluation methodology and suite of metrics. SEAN can be easily extended with new metrics.

1.2 Gathering Human Feedback

Human feedback for social navigation algorithms is often crowdsourced via survey questions after having participants watch a video of the robot navigating. We hypothesized that the perception of a robot would differ between a participant that watched a video of a human and robot interacting versus a participant that directly interacted with the robot in simulation. Previously, the interaction setup would have been challenging to test at scale. However, we recently studied how SEAN can be deployed at scale on the web [17]. Using modern web technologies, we embedded SEAN in a web-based survey form and quickly collected data from 62 participants about robot navigation. They controlled a virtual character in the simulator from a third person camera view that followed the avatar. They performed simple tasks that encouraged them to interact with a simulated robot controlled via ROS. For example, we asked participants to find and follow the robot. Participants then rated the robot along the Competence and Discomfort factors of the Robot Social Attribute Scale (ROSAS) [7].

We also collected data for another 62 participants that watched video recordings of the previous participants' interactions from the same camera perspective. Video participants also rated the robot's Competence and Discomfort [7]. Using a Wilcoxon signed-rank test on pairs of interactive and video data we found no significant difference for the Discomfort factor of ROSAS. However, the Wilcoxon test revealed significant differences for the Competence factor ($p<0.0001$). The median Competence value was 3 for the interactive condition and 3.83 for the video condition. We suspect the different results were due to the simulation's interactive nature.

We plan to perform an in-person study to verify that simulations lead to more realistic evaluations than video surveys. Several of the interactions considered in the prior study took place in a virtual model of our lab (Fig. 1a). Our in-person study will recreate these interactions using our real world lab and real Kuri robot.

1.3 Understanding and Responding to Failure

Now able to collect human feedback at scale, our focus has shifted to exploring methods that can fundamentally advance the reliability of social navigation. In a small step towards bridging the gap between teleoperation and full autonomy, we are leveraging the simulation platform to develop computational methods for detecting robot failures. A method for detecting and responding to errors in a reasonable manner will improve the reliability of social navigation robots while minimizing the need for realtime human supervision.

Data from human supervision will be collected via the simulation platform deployed at scale. Crowdworkers will be asked to continuously predict if a simulated robot will successfully complete a navigation behavior between indicated points. Crowdworkers will also annotate points in time that the robot appears to make unsafe or unpredictable motions. Further, these unsafe movements will be labeled based on crowdworkers' concerns, e.g., the robot was unnecessarily too close to a human, or it moved too fast nearby some static object. After each interaction, crowdworkers will answer survey questions about the robot's motion. For example, questions will cover perceived safety, predictability, and legibility.

We plan to collect implicit feedback by extending our crowdsourcing infrastructure to record webcamera videos of participants faces as they experience the simulation. We will automatically extract facial Action Units [10] from these videos as we have done previously [8] via [6] and correlate them to the explicit feedback.

Building on prior experience in computer vision [14] and classification [18], we will study neural-network based multi-class classifiers to predict unsafe robot movements. The key question under investigation is whether classifiers conditioned on implicit social signals can reach the efficacy of those conditioned on explicit data. If so, in the future, implicit annotations such as facial expressions of nearby people could be used to detect unsafe motions on deployed social robotic platforms. We also plan to apply the same method to detection of social norm violations in specific social settings.

2 CONCLUSION

We hope that our work towards benchmarking autonomous social navigation will help advance community dialog around evaluation metrics and methodologies. SEAN deployed on the web can enable data collection for training new navigation methods as well as facilitate the collection of human feedback at a scale not commonly seen in HRI. Currently, we are working towards building methods for more reliable social navigation by leveraging implicit and explicit human feedback. We expect this type of approach to bring social robotics a step closer to the reliability levels that real users demand.

3 ACKNOWLEDGEMENTS

Thanks to M. Hussein, O. Fugikawa, JD Zhao, J. Espinoza, and X. Ruiz for helping develop SEAN. Thanks also to Amazon and the National Science Foundation (IIS-1924802) for supporting this work.

REFERENCES

[1] 1990. IEEE Standard Glossary of Software Engineering Terminology. *IEEE Std 610.12-1990* (1990), 1–84. https://doi.org/10.1109/IEEESTD.1990.101064

[2] 2020. ROS# open source software libraries and tools. https://github.com/siemens/ros-sharp. Accessed: 2020-11-11.

[3] 2020. Rosbridge provides a JSON API to ROS functionality for non-ROS programs. http://wiki.ros.org/rosbridge_suite. Accessed: 2020-11-11.

[4] 2020. SEAN documentation and open source code. https://sean.interactive-machines.com. Accessed: 2020-11-11.

[5] 2020. Unity Real-Time Development Platform. https://www.unity.com/. Accessed: 2020-11-11.

[6] Tadas Baltrusaitis, Amir Zadeh, Yao Chong Lim, and Louis-Philippe Morency. 2018. Openface 2.0: Facial behavior analysis toolkit. In *2018 13th IEEE International Conference on Automatic Face & Gesture Recognition (FG 2018)*. IEEE, 59–66.

[7] Colleen M Carpinella, Alisa B Wyman, Michael A Perez, and Steven J Stroessner. 2017. The robotic social attributes scale (RoSAS) development and validation. In *Proceedings of the 2017 ACM/IEEE International Conference on human-robot interaction*. 254–262.

[8] Joe Connolly, Viola Mocz, Nicole Salomons, Joseph Valdez, Nathan Tsoi, Brian Scassellati, and Marynel Vázquez. 2020. Prompting Prosocial Human Interventions in Response to Robot Mistreatment. In *ACM/IEEE International Conference on Human-Robot Interaction (HRI)*.

[9] Joe Connolly, Nathan Tsoi, and Marynel Vázquez. 2021. Perceptions of Conversational Group Membership based on Robots’ Spatial Positioning: Effects of Embodiment. In *Companion of the 2021 ACM/IEEE International Conference on Human-Robot Interaction*.

[10] Paul Ekman. 1977. Facial action coding system. (1977).

[11] Anna Konrad. 2019. *Simulation of Mobile Robots with Unity and ROS: A Case-Study and a Comparison with Gazebo*. Master’s thesis. Department of Engineering Science, University West.

[12] Evan Marcus and Hal Stern. 2003. *Blueprints for high availability*. John Wiley & Sons.

[13] Morgan Quigley, Ken Conley, Brian Gerkey, Josh Faust, Tully Foote, Jeremy Leibs, Rob Wheeler, and Andrew Y Ng. 2009. ROS: an open-source Robot Operating System. In *ICRA workshop on open source software*, Vol. 3. Kobe, Japan, 5.

[14] Hamid Rezatofighi, Nathan Tsoi, JunYoung Gwak, Amir Sadeghian, Ian Reid, and Silvio Savarese. 2019. Generalized Intersection over Union. In *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*.

[15] Samuel S. Sohn, Honglu Zhou, Seonghyeon Moon, Sejong Yoon, Vladimir Pavlovic, , and Mubbasir Kapadia. 2020. Laying the Foundations of Deep Long-Term Crowd Flow Prediction. In *Proceedings of the European Conference on Computer Vision (ECCV)*.

[16] Nathan Tsoi, Mohamed Hussein, Jeacy Espinoza, Xavier Ruiz, and Marynel Vázquez. 2020. SEAN: Social Environment for Autonomous Navigation. In *Proceedings of the 8th International Conference on Human-Agent Interaction (HAI)*.

[17] Nathan Tsoi, Mohamed Hussein, Olivia Fugikawa, JD Zhao, and Marynel Vázquez. 2020. SEAN-EP: A Platform for Collecting Human Feedback for Social Robot Navigation at Scale. arXiv:2012.12336 [cs.RO]

[18] Nathan Tsoi, Yofti Milkessa, and Marynel Vázquez. 2020. A Heaviside Function Approximation for Neural Network Binary Classification. arXiv:2009.01367 [cs.LG] https://arxiv.org/abs/2009.01367

[19] Gloria Zen, Bruno Lepri, Elisa Ricci, and Oswald Lanz. 2010. Space speaks: towards socially and personality aware visual surveillance. In *Proceedings of the 1st ACM international workshop on Multimodal pervasive video analysis*. 37–42.