Metrics for Evaluating Social Conformity of Crowd Navigation Algorithms

Junxian Wang[1], Wesley P. Chan[*,1], Pamela Carreno-Medrano[1], Akansel Cosgun[1], Elizabeth Croft[2]

I. Introduction

Autonomous robot navigation in populated environments has been an active research field in the past decade [1], [2]. A key challenge is to design algorithms that allow robots to navigate safely and socially in such environments. Recently, advancements in computation hardware and machine learning algorithms have enabled a series of deep learning based methods to emerge [3]–[6]. These works have diversified definitions of what constitutes "social behavior", as well as evaluation criteria for social conformity of the resulting robot navigation behavior. Given this lack of consistency in social evaluation metrics used for crowd navigation algorithms, it is not surprising to find that publications that aim to provide socially conforming crowd navigation algorithms usually present evaluations that are mainly focused on efficiency, and lack details on the social conformity of the trained results. Furthermore, the lack of a set of well defined standardized metrics make it difficult, if not impossible, to compare performances of algorithms published in different works.

To fill this gap, this work proposes a set of metrics intended to be used for evaluating and comparing different crowd navigation algorithms from a social conformity aspect. The proposed metrics are applied to a collection of state-of-the-art crowd navigation algorithms.

II. Proposed Metrics

Literature defines three categories of social navigation criteria [7]: comfort, naturalness and sociability. We propose the following evaluation metrics, covering all categories:

- **Comfort:** the absence of annoyance and stress for humans in interaction with robots
  - (I) Duration of robot being inside the minimum accepted personal space of strangers
  - (II) Duration of when robot's immediate projected path intersections with others' projected path
  - (III) Aggregated goal reaching time for all cooperative agents
- **Naturalness:** the similarity between robots and humans in low-level behavior patterns
  - (IV) Integrated jerk over the robot's trajectory (c.f. Minimum Jerk Trajectory humans employ)
  - (V) Duration of when robot's speed exceeds human maximum normal walking speed
- **Sociability:** the adherence to explicit high-level cultural conventions
  - (VI) Side preference (either left or right) percentage for passing / overtaking / crossing behavior

[1]Department of Electrical and Computer Systems Engineering, Monash University

[2]Faculty of Engineering, Monash University

[*]Corresponding author

III. Evaluation Method

Using the proposed metric set, we evaluate in simulation four recently published algorithms that have made their implementation publicly available, namely CADRL [3], LSTM-RL [4], SARL [5], RGL [6]. We trained the selected algorithms on a same CrowdSim [6] environment with similar configurations. The training parameters are uniform across all algorithms to ensure training is done under similar conditions. Algorithms are trained for 10000 episodes using a rotation of 4 selected scenarios.

The algorithms are tested and evaluated using our proposed metrics on a set of 7 simulated scenarios: circular crossing, random spawn, parallel traffic, perpendicular traffic, and one scenario each for passing, overtaking and crossing other humans, each with 5 simulated pedestrians. Pedestrians have variable preferred speeds and react to the robot's motion using a reactive collision avoidance model, ORCA [8].

IV. Preliminary Results

Table I shows initial partial evaluation results of the algorithms using a few of the proposed metrics in one scenario. Results show averages calculated from successful cases within 500 test cases.

Algorithm	Evaluation Metric			
	I (s)	**II** (s)	**III** (s)	**IV** (m/s^2)
CADRL [3]	36.2 ± 16.9	7.0 ± 5.8	243.5 ± 33.7	105.0 ± 12.4
LSTM-RL [4]	**25.1** ± 14.4	**3.2** ± 4.9	278.0 ± 48.4	115.7 ± 14.7
SARL [5]	29.7 ± 11.6	4.5 ± 4.2	260.7 ± 35.3	65.3 ± 36.3
RGL [6]	25.4 ± 12.7	6.7 ± 4.9	**242.2** ± 21.2	**22.7** ± 10.4

TABLE I: Test results of chosen algorithms. Bold fonts show best performance. Metrics V and VI are being evaluated in our ongoing work.

References

[1] A. Cosgun, *et al.*, "Anticipatory robot path planning in human environments," in *RO-MAN*, 2016.

[2] Y. Du, *et al.*, "Group surfing: A pedestrian-based approach to sidewalk robot navigation," in *ICRA*, 2019.

[3] Y. F. Chen, *et al.*, "Decentralized non-communicating multiagent collision avoidance with deep reinforcement learning," in *ICRA*, 2017.

[4] M. Everett, *et al.*, "Motion planning among dynamic, decision-making agents with deep reinforcement learning," in *IROS*, 2018.

[5] C. Changan, *et al.*, "Crowd-Robot Interaction: Crowd-aware Robot Navigation with Attention-based Deep Reinforcement Learning," *arXiv:1809.08835 [cs]*, 2019.

[6] C. Changan, *et al.*, "Relational Graph Learning for Crowd Navigation," *arXiv:1909.13165 [cs]*, 2020.

[7] T. Kruse, *et al.*, "Human-aware robot navigation: A survey," *Robotics and Autonomous Systems*, vol. 61, no. 12, pp. 1726–1743, 2013.

[8] J. Van Den Berg, *et al.*, "Reciprocal n-body collision avoidance," in *Robotics research*. Springer, 2011, pp. 3–19.