Legibot: Legible Motions for Service Robots

Javad Amirian

January 9, 2024

Abstract

With the prevalence of social robots in various environments and applications, there is an increasing need for these robots to exhibit socially-compliant behaviors. Legible motion, characterized by the ability of a robot to clearly and quickly convey intentions and goals to the individuals in its vicinity, holds significant importance in this context. In this paper, we introduce a novel approach to incorporate legibility into local motion planning for mobile robots. This can enable robots to generate legible motions in real-time and dynamic environments. To demonstrate the effectiveness of our proposed methodology, we conducted real-world experiments involving the Pepper robot and simulated scenarios featuring mobile robots in a restaurant environment with multiple human occupants.

1 Introduction

Robotic systems have transcended their traditional roles in factories and manufacturing lines, expanding into various service-oriented domains, including healthcare, hospitality, and food service. As robots increasingly share spaces with humans, it becomes imperative for them to comprehend and adhere to the implicit social norms that govern human interaction. This imperative gives rise to the concept of social-compliant behaviors, wherein robots are expected to exhibit behaviors that align with human expectations. Francis et al. [7] have identified eight distinct sets of principles that collectively define social compliance for robots, encompassing aspects such as safety, politeness, and legibility. Notably, these dimensions are not mutually exclusive; improvements in one dimension can influence others, either positively or negatively.

Among these dimensions, legibility remains an underdeveloped aspect in the field of robotics and presents substantial room for advancement. Legibility in robotic motion refers to the robot's capacity to clearly and swiftly communicate its intentions and objectives to individuals in its vicinity. In the context of human-robot interaction, achieving legible motion is of paramount significance, as it enhances user understanding, trust, and overall user experience. It might sometimes be simply as the agent's effort to exaggerate its action to make sure the opponent is aware of its decision, which can be critical in human-robot tasks that require tight collaboration between the two parties. But, also it can appear in more complex scenarios to respect the social rules in a certain space, and adapt to some acceptable behaviors in that context.

Contributions: In this paper we propose a novel approach to incorporate legibility into local motion planning for mobile robots. The proposed formulation enables robots to generate legible motions in real-time and dynamic environments. These dynamic changes in the environment can appear in different forms, such as a moving target, finding about new irrelevant goals, and also the set of observers that robot should be legible for. ...

Structure: In the next section, we review the related work in the field **??**. Then in section **??** we propose the aforementioned framework in detail, and then in section **??** we explain the experiments we conducted with simulated and real robots to demonstrate the effectiveness of our proposed methodology in a restaurant scenario.

2 Related Work

In [4] Dragan et al. differentiate predictability and legibility, crucial for human-robot collaboration. They provide formal definitions, propose cost-based models for motion planning, and practically validate the contradiction between predictability and legibility in various characters.

Work [9] addresses the problem of the limited field of view of observers. Their proposed algorithm models observer locations and perspectives, enhancing legibility by placing movements where easily be seen. The study shows that observer-aware legibility increases the duration of correct goal inferences, but non-targeted observers have lower performance when paths are personalized for others. The paper emphasizes the importance of considering an observer's environment for effective planning in scenarios like robot-assisted **restaurant** service. Also [3] implemented a **coffee-shop** scenario. Participants collaborated with the robot to fulfill tea orders. The robot retrieved the appropriate cup, and participants selected ingredients based on the cup being retrieved.

In [1] the core idea is bringing end-to-end framework using conditional generative models to learn legible robot trajectories from multi-modal human demonstrations.

* Of course, we all are aware of how signaling lights are used in vehicles to show other people our intentions, there are also other modalities to communicate the intent can be used, for example, in IAN by Dugas et al. [5] a Pepper robot uses hand gestures and nudges to communicate with the people around it and ask them to clear the way.

* [8]

Learning-based approaches have also been used to generate legible motions. In SLOT-V [**?**] the authors present a method that takes labeled robot trajectories and learns an observer model, using a deep neural network as a value function approximator. However, assuming that the robot has access to a labeled dataset of legible trajectories is not always realistic.

3 Background

3.1 Legibility Score

The original formula for motion legibility (by Dragan et al. [4]) for an observed trajectory ξ [4] is as follows:

$$\text{legibility }(\xi)=\frac{\int P\left(G^{*} \mid \xi_{S \rightarrow \xi(t)}\right) f(t) d t}{\int f(t) d t} \tag{1}$$

This equation assumes $P\left(G \mid \xi_{S \rightarrow \xi(t)}\right)$ is the way an observer distributes probabilities to potential goals of the robot, with G^* being the true goal of the robot, and $f(t)$ to be a descending function like $f(t) = T - t$ which assigns higher weights to the initial parts of the trajectory, justifying the fact that a legible motion should minimize the ambiguity as soon as possible for the observers.
To compute Eq. (1), they use the Bayse' rule and rewrite $P\left(G^{*} \mid \xi_{S \rightarrow \xi(t)}\right)$ as below:

$$P\left(G \mid \xi_{S \rightarrow Q}\right) \propto \frac{\exp \left(-C\left(\xi_{S \rightarrow Q}\right)-C\left(\xi_{Q \rightarrow G}^{*}\right)\right)}{\exp \left(-C\left(\xi_{S \rightarrow G}^{*}\right)\right)} P(G) \tag{2}$$

with $P(G)$ being a prior distribution of the potential goals and $C(\xi)$ being a cost function for an observed trajectory ξ or an optimal trajectory ξ^*.

3.2 Synthesis

Here we explaing our contributions, a new algorithm to generate legible motions for mobile robots. In fact, the problem of generating legible motions for mobile robots has not received as much attention as the evaluation. For example, in several works legible and illegible trajectories were manually designed by the authors, or handcrafted to collect the user study data [REFs]. And in [Ada-RO-MAN2022] the paths were selected via a sampling approach.

From [X] we know that optimizing the legibility function directly is not always possible. This is because that $\Delta\mathbf{L}(\xi) = 0$ or $\mathbf{L}(\xi) = 1$ might not have a solution in the finite space. For this reason, we might need to add some constraints to the equation to make it solvable. In the original work, Dragana et al. [X] this is done by adding a regularizer that discourage increasing the path length. $L(\xi) = \mathit{legibility}(\xi) - \lambda C(\xi)$ where $C(\xi)$ is the path length, and λ is a constant. Also, Dragan and Srinivasa [X], introduce a trust region constraint on the optimization to ensure the motion does not become too surprising or unpredictable to the observer. This approach in the end, turns to finding a good value for a parameter β, using a user study. Moreover, due to the iterative nature of this approach, it can compromise the real-time performance of the system.

In this work, we propose to use local planning to generate legible motions for mobile robots. We show that this approach can improve the time performance of the system, and can generate legible motions in real-time. Our formulation takes into account the final pose of the robot at the end of the motion, and introduces a new cost function to the local planner, to generate legible motions.

We also show how this new formulation makes it more stable and easy to **dynamically** consider new observers in the environment, once the robot has detected them, or to consider the change in the observer's field of view, by simply re-planning the motion.

4 Synthesis

The main motivation of this part is to propose a practical and simple algorithm to generate legible motions for mobile robots. First of all, synthesizing legible behaviors can be explained as "finding behaviors that convince one or multiple observer that the robot is doing what it is supposed to do". This is beyond the evaluation problem, where we are trying to find a metric to evaluate the legibility of a given behavior [2]. And, we would like to highlight that the synthesis problem has not received as much attention as the evaluation in the literature, mainly because of the complexity of the problem.

We assume our mobile robot has a goal of reaching a target G^* that might move over time. However, there are other goals G^i that are irrelevant to the robot's task, all forming a set of goals $\mathcal{G}$. In the environment, there will be a set of observers $\mathcal{O}$, that contains a positive number of observers O^j that are interested in the robot's behavior. The robot is supposed to reach the goal G^* in a way that is legible for all the observers O^j. Note that, the observers set $\mathcal{O}$ and the goals set $\mathcal{G}$ might be the same, but not necessarily. For example, in a restaurant scenario, the robot might want to deliver a dish to a customer (G^*), while there might be other customers in the restaurant $\mathcal{G} \neq \mathrm{G}^*$, and we assume that all of them are interested in the robot's behavior ($\mathcal{O} = \mathcal{G}$). Finally, the robot is moving in an environment $\mathcal{E}$, with an obstacle-free space $\mathcal{E}^t_{free}$ that might also change over time. We are interested in finding a path $\xi_{t_0:t_w}$ from the current time t_0 to the time t_w that is legible for all the observers O^j, while satisfying the robot's task, i.e. reaching the goal G^*, avoiding the static and dynamic obstacles, moving within the robot's kinematic and dynamic constraints, and so on.

In this work, we propose to use local planning to generate legible motions. Back to local planning algorithms, such as Dynamic Window Approach (DWA) [6], are designed to generate a motion that is feasible and optimal in terms of a cost function. In this family of algorithms, instead of finding a path from the start to the goal, the robot is trying to find a motion that is optimal in terms of a cost function, over a short horizon, or a Window. This window, denoted by w, is defined in such a way that the robot can stop before hitting an obstacle, by incorporating the robot's kinematic and dynamic constraints. But also, to ensure the real-time performance of the algorithm, in environments where a full plan is too expensive to compute.

Let $\xi_{t_0:t_0+w}$ be the path we want to plan, from the current time t_0 over a window w. We define a cost function $C(\xi)$ to optimize over this window. This cost function can be in the form of a weighted sum of different terms, such as the distance to the goal, the distance to the obstacles, and the smoothness of the path:

$$C(\xi_{t_0:t_0+w}) = \sum_i \alpha_i C^i_{task}(\xi_{t_0:t_0+w}) \tag{3}$$

where $C^i_{task}(.)$ is the cost of the i-th term, and α_i is the weight of the i-th term. We propose an extra term to this cost function, to incorporate the legibility of the motion, and we call it $C^j_I(\xi_{t_0:t_0+w})$, representing the illegibility of the motion for the j-th observer.

We assume the trajectory $\xi_{t_0:t_0+w}$ is a sequence of sub-trajectories $\xi_{t:t+dt}$, or simply velocity vector $\mathbf{v}_t$, where dt is the time step. Hence, a sub-trajectory $\xi_{t:t+dt}$ incurs an extra cost of $\Delta C^j_I(\xi_{t:t+dt})$ for the j-th observer. We assume that this cost is proportional to the probability of the observer not being able to predict the true intention of the robot, i.e.:

$$\Delta C_I^j(\xi_{t:t+dt}) \propto \mathrm{P}^j(G \neq G^*|\mathbf{v}_t)\ dt \tag{4}$$

Here, instead of assuming the observer has access to the full path $\xi_{t_0:t+dt}$, we assume the observer only has access to the current state of the robot $\xi_{t:t+dt}$. This helps us to, first, to make the cost function more computationally efficient. and second, to make it more realistic, as we have no information if the robot has been in the observer's field of view in the past, or if the observer has actively been paying attention to the robot's behavior.

$$\mathrm{P}^j(G \neq G^*|\mathbf{v}_t) = \sum_{G \neq G^*} \mathrm{P}^j(G|\mathbf{v}_t) \tag{5}$$

We can approximate $\mathrm{P}^j(G|\mathbf{v}_t)$ with the projected cosine distance between $\mathbf{v}_t$ and the optimal path from $\mathbf{x}_t$ to $G \neq G^*$:

$$\mathrm{P}(G \neq \mathrm{G}^*|\mathbf{v}_t) \approx \sum_{j \in \mathcal{O}} \sum_{G \neq \mathrm{G}^*} d\measuredangle(\mathrm{T}^j \mathbf{v}_t, \mathrm{T}^j \mathbf{v}_t^*(.|G)) \tag{6}$$

where T^j is the transformation matrix from the world frame to the observer's frame j, and $\mathbf{v}_{t:t+dt}(.|\mathrm{G}^*)$ is the velocity vector of the robot at time t, Here we compute the optimal paths also by solving Eq. (3) for each goal G, and then we compute the cosine distance between the velocity vector of the robot and the optimal velocity vector for each goal G.

5 Extra Points and Comments

* Note that the "shortcut heuristic" used in traditional planners, such as DWA, is not applicable here, since finding a shortcut even though might reduce other costs of the plan, it can increase the legibility cost.

Extra: In fact, optimizing the legibility function directly is not always possible [REF]. This is because that $\Delta\mathbf{L}(\xi) = 0$ or $\mathbf{L}(\xi) = 1$ might not have a solution in the finite space. For this reason, we might need to add some constraints to the equation to make it solvable. In the original work, Dragana et al. [X] this is done by adding a regularizer that discourage increasing the path length. $L(\xi) = legibility(\xi) - \lambda C(\xi)$ where $C(\xi)$ is the path length, and λ is a constant. Moreover, due to the iterative nature of this approach, it can compromise the real-time performance of the system.

6 Experimental Results

6.1 Simulation in Restaurant Setup

One challenge with computing Optical Flow on simulated frames is that due to the lack of texture, it might become unstable during frames where there is zero or low level of motion in the scene. To fix this we have to preprocess the frames with a simple subtraction operator between the consecutive frames and estimate the total motion before passing the image to the optical-flow deep network.

6.2 Why restaurants?

In fact, robots have been already deployed in restaurants and thousands of them are in place. This type of situation usually contains a certain number of people: clients, waiters, ... And the people are part of interaction scenarios with the robot(s). The clients might be new to this place and new to facing the robot in this restaurant, which means they wouldn't have a clear idea of the robot's behavior.

6.3 Experiments with Real Robot

6.4 User Study

7 Conclusion

References

[1] Matthew Bronars and Danfei Xu. Legible robot motion from conditional generative models. 2023.

[2] Anca Dragan and Siddhartha Srinivasa. Generating legible motion.

[3] Anca D Dragan, Shira Bauman, Jodi Forlizzi, and Siddhartha S Srinivasa. Effects of robot motion on human-robot collaboration. In *Proceedings of the Tenth Annual ACM/IEEE International Conference on Human-Robot Interaction*, pages 51–58, 2015.

[4] Anca D Dragan, Kenton CT Lee, and Siddhartha S Srinivasa. Legibility and predictability of robot motion. In *2013 8th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, pages 301–308. IEEE, 2013.

[5] Daniel Dugas, Juan Nieto, Roland Siegwart, and Jen Jen Chung. Ian: Multi-behavior navigation planning for robots in real, crowded environments. In *2020 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pages 11368–11375. IEEE, 2020.

[6] Dieter Fox, Wolfram Burgard, and Sebastian Thrun. The dynamic window approach to collision avoidance. *IEEE Robotics & Automation Magazine*, 4(1):23–33, 1997.

[7] Anthony Francis, Claudia Pérez-d'Arpino, Chengshu Li, Fei Xia, Alexandre Alahi, Rachid Alami, Aniket Bera, Abhijat Biswas, Joydeep Biswas, Rohan Chandra, et al. Principles and guidelines for evaluating social robot navigation algorithms. *arXiv preprint arXiv:2306.16740*, 2023.

[8] Masaki Fukuchi, Naotsugu Tsuchiya, and Christof Koch. The focus of expansion in optical flow fields acts as a strong cue for visual attention. *Journal of Vision*, 9(8):137–137, 2009.

[9] Ada V Taylor, Ellie Mamantov, and Henny Admoni. Observer-aware legibility for social navigation. In *2022 31st IEEE International Conference on Robot and Human Interactive Communication (RO-MAN)*, pages 1115–1122. IEEE, 2022.