Explainable Robot Navigation

Amar Halilovic

Institute of Artificial Intelligence, Ulm University
Ulm, Germany
amar.halilovic@uni-ulm.de

Abstract

As the use of autonomous mobile robots expands into dynamic and complex environments, the need for them to provide understandable explanations for their actions becomes crucial. This thesis addresses the challenge of developing explainability for robot navigation by leveraging a hybrid model that combines machine learning techniques with symbolic reasoning methods. Furthermore, the thesis explores the modeling of human explanation preferences and the impact of different explanation attributes on explanation recipients' understanding, satisfaction, and trust. The goal is to integrate different explanation aspects and approaches into a unified framework to support explainable navigation in robotics.

Introduction

With the growing adoption of autonomous systems, including mobile robots in environments like factories, hospitals, and homes, their navigation and interaction complexity has significantly increased. Service robots, whose primary role is to provide assistance or perform tasks for humans, face unique challenges in navigating dynamic settings while adhering to social norms and user expectations. Although robotics and artificial intelligence (AI) advancements have led to impressive developments in autonomous decision-making, current systems often fail to provide explanations that align with human reasoning processes, creating a gap between robotic perception and human expectations.

Explainable Artificial Intelligence (XAI) seeks to bridge this gap by making autonomous systems transparent and understandable to humans. In the context of robot navigation, the ability to explain actions is crucial for building trust and ensuring safe and effective human-robot interaction (HRI). This research aims to develop a comprehensive hierarchical framework for explainable robot navigation, leveraging a hybrid model that integrates machine learning with symbolic reasoning on different abstraction levels, i.e., lower path-planning and higher task-planning abstraction levels. The main research questions investigated are:

- Can we develop and/or equip navigation algorithms with capabilities that enable robots to generate understandable explanations of their actions, improving HRI?
- What explanations do users prefer based on explanation attributes such as type, modality, and timing?
- Are explanations at higher or lower levels of abstraction in robot navigation more effective for users?
- Can different abstraction levels in explanation planning for robot navigation be integrated into a framework?

Related Work

Making robot navigation explainable is part of the bigger goal of Explainable AI: making complex algorithms more transparent and helping to ensure robots are used safely and morally. So far, studies on explainable robot navigation have mainly examined why robots fail at path planning (Kwon, Huang, and Dragan 2018) or choose paths, focusing on comparing different paths (Krarup et al. 2021). Explanations in robotics are usually presented visually (He, Aouf, and Song 2021) and textually (in natural language) (Rosenthal, Selvaraj, and Veloso 2016). Several studies have found that giving explanations can help users understand autonomous systems better (Kwon, Huang, and Dragan 2018) and trust them more (Stange and Kopp 2020). The literature on navigation task planning explanations usually differs between explicable planning (Zhang et al. 2017) and plan explanation as model reconciliation (Chakraborti et al. 2017). Explicable planning means that robots try to produce inherently interpretable plans that are close to human expectations. On the other hand, the latter approach aims to make robots' original plans more explainable to humans by updating human expectations through explanations.

Research Plan and Methodology

This research employs machine learning and automated planning (AI planning) to enhance the explainability of robot navigation. The approach is twofold:

1. **Path Planning (Lower Abstraction Level)**: I explore the use of affordance theory from psychology to understand how different objects in the robot's environment afford specific actions. Using machine learning techniques, I teach the robot to recognize objects' actionable properties (affordances) and how to act upon them. For example, the robot can identify that a door affords the action of opening. The knowledge derived is then used to generate

visual and textual explanations that illustrate why certain objects are treated as obstacles or targets for interaction.

2. **Task Planning (Higher Abstraction Level)**: Task planning is modeled using AI Planning, which involves defining high-level tasks such as "go to the kitchen" or "bring me a book." Explanation generation is also modeled as an automated planning problem, where explanations are actions that robots can plan together with other non-explanation-related actions. The explanation content is based on the lower level. For instance, when the robot encounters an obstacle, it can generate an explanation action suggesting humans remove the object, backed by its understanding of the object's affordances.

Results

The current results indicate that simple XAI models, specifically inherently explainable models, can be used to provide knowledge to a robot regarding the extent to which each environmental obstacle contributes to the robot's state. The robot can then use this knowledge to create visual explanations (Halilovic and Lindner 2022). Initially, my focus was limited to the ability to move objects. In my second paper (Halilovic and Lindner 2023), I introduced multiple affordances and showed how robot explanations can be verbalized in the form of textual explanations that accompany visual explanations, thereby creating visual-textual explanation coherence. The affordances of objects are now stored in a database, which enables the robot to retrieve the necessary knowledge as needed. Subsequently, I explored a higher level of abstraction, presenting a method for how a robot can generate structured explanations (Halilovic and Krivic 2023b). I also examined how a robot's internal state, such as human-like traits (e.g., extroversion), can affect the generation of explanations (Halilovic and Krivic 2023a). The next study involved a more extensive user study, where we explored which modalities of explanations people prefer—visual, textual, or visual-textual. The results showed that, in most situations, people prefer visual-textual explanations (Halilovic, Chandrayan, and Krivic 2024), as they provide the most information. The next step involved modeling the process of generating explanations as a discrete sequential problem (Halilovic and Krivic 2024). More recently, I have worked on probabilistic explanation generation (Halilovic, Krivić, and Canal 2024) (extended paper submitted to AAAI-25), where the robot attempts to respect human preferences when generating explanations.

Future Work

By the workshop date (February 25-26, 2025), I anticipate the following progress:

- I plan to conduct further user studies to determine how other explanation attributes (detail level, timing, type) affect explanation recipients' understanding, satisfaction, and trust.
- I plan on developing the first working version of a unified hierarchical framework that integrates different levels of explanation abstraction.

References

Chakraborti, T.; Sreedharan, S.; Zhang, Y.; and Kambhampati, S. 2017. Plan explanations as model reconciliation: Moving beyond explanation as soliloquy. *arXiv preprint arXiv:1701.08317*.

Halilovic, A.; Chandrayan, V.; and Krivic, S. 2024. Exploring the Impact of Explanation Representation on User Satisfaction in Robot Navigation. In *Proceedings of the 2024 International Symposium on Technological Advances in Human-Robot Interaction*, 1–9.

Halilovic, A.; and Krivic, S. 2023a. The influence of a robot's personality on real-time explanations of its navigation. In *International Conference on Social Robotics*, 133–147. Springer.

Halilovic, A.; and Krivic, S. 2023b. Towards a Holistic Framework for Explainable Robot Navigation. In *International Workshop on Human-Friendly Robotics*, 213–228. Springer.

Halilovic, A.; and Krivic, S. 2024. Planning of explanations for robot navigation. In *2024 IEEE International Conference on Robotics and Automation (ICRA)*, 5478–5484. IEEE.

Halilovic, A.; Krivić, S.; and Canal, G. 2024. Towards Probabilistic Planning of Explanations for Robot Navigation. In *RSS 2024 Workshop on Unsolved Problems in Social Robot Navigation*.

Halilovic, A.; and Lindner, F. 2022. Explaining local path plans using LIME. In *International Conference on Robotics in Alpe-Adria Danube Region*, 106–113. Springer.

Halilovic, A.; and Lindner, F. 2023. Visuo-Textual Explanations of a Robot's Navigational Choices. In *Companion of the 2023 ACM/IEEE International Conference on Human-Robot Interaction*, 531–535.

He, L.; Aouf, N.; and Song, B. 2021. Explainable Deep Reinforcement Learning for UAV autonomous path planning. *Aerospace Science and Technology*, 118: 107052.

Krarup, B.; Krivic, S.; Magazzeni, D.; Long, D.; Cashmore, M.; and Smith, D. E. 2021. Contrastive explanations of plans through model restrictions. *Journal of Artificial Intelligence Research*, 72: 533–612.

Kwon, M.; Huang, S. H.; and Dragan, A. D. 2018. Expressing robot incapability. In *Proceedings of the 2018 ACM/IEEE International Conference on Human-Robot Interaction*, 87–95.

Rosenthal, S.; Selvaraj, S. P.; and Veloso, M. M. 2016. Verbalization: Narration of Autonomous Robot Experience. In *IJCAI*, volume 16, 862–868.

Stange, S.; and Kopp, S. 2020. Effects of a social robot's self-explanations on how humans understand and evaluate its behavior. In *Proceedings of the 2020 ACM/IEEE international conference on human-robot interaction*, 619–627.

Zhang, Y.; Sreedharan, S.; Kulkarni, A.; Chakraborti, T.; Zhuo, H. H.; and Kambhampati, S. 2017. Plan explicability and predictability for robot task planning. In *2017 IEEE international conference on robotics and automation (ICRA)*, 1313–1320. IEEE.