Self-Explainable Robots in Remote Environments

Francisco J. Chiyah Garcia*
Heriot-Watt University
Edinburgh, United Kingdom
fjc3@hw.ac.uk

Simón C. Smith*
The University of Edinburgh
Edinburgh, United Kingdom
artificialsimon@ed.ac.uk

José Lopes
Heriot-Watt University
Edinburgh, United Kingdom
jd.lopes@hw.ac.uk

Subramanian Ramamoorthy
The University of Edinburgh
Edinburgh, United Kingdom
s.ramamoorthy@ed.ac.uk

Helen Hastie
Heriot-Watt University
Edinburgh, United Kingdom
h.hastie@hw.ac.uk

ABSTRACT

As robots and autonomous systems become more adept at handling complex scenarios, their underlying mechanisms also become increasingly complex and opaque. This lack of transparency can give rise to unverifiable behaviours, limiting the use of robots in a number of applications including high-stakes scenarios, e.g. self-driving cars or first responders. In this paper and accompanying video, we present a system that learns from demonstrations to inspect areas in a remote environment and to explain robot behaviour. Using semi-supervised learning, the robot is able to inspect an offshore platform autonomously, whilst explaining its decision process both through both image-based and natural language-based interfaces.

KEYWORDS

Explainable robot, semi-supervised learning, autonomous control, transparent interfaces, remote location, NLG

ACM Reference Format:

Francisco J. Chiyah Garcia, Simón C. Smith, José Lopes, Subramanian Ramamoorthy, and Helen Hastie. 2021. Self-Explainable Robots in Remote Environments. 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/10.1145/3434074.3447275

1 INTRODUCTION

Recent advancements have made robots and autonomous systems a valuable asset in unsafe environments as a way to keep humans out of danger, such as in the nuclear or energy domains [6, 7, 10, 13, 17, 20]. These domains often require navigating highly dynamic scenarios and executing time-critical actions successfully. We focus here on offshore energy platforms as part of the EPSRC ORCA Hub programme [6], where robots are expected to maintain and operate parts of the platform autonomously, whilst communicating with remote operators through a human-robot interface.

*Both authors contributed equally to this research.

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

ACM ISBN 978-1-4503-8290-8/21/03.
https://doi.org/10.1145/3434074.3447275

Machine learning is an important part of modern robots, with more accurate and sophisticated algorithms being deployed over time. However, it requires a high amount of data, which is challenging and costly to obtain, particularly in hazardous domains. Using semi-supervised machine learning, we have shown that a robot can learn inspection strategies from a demonstration by a human operator [19]. This method provides several advantages over both supervised and unsupervised machine learning algorithms: it does not require expensive data collections and it leverages unsupervised learning methods to extract information not explicitly available. More importantly, the robot successfully adapts to unseen conditions, which is imperative in changing situations.

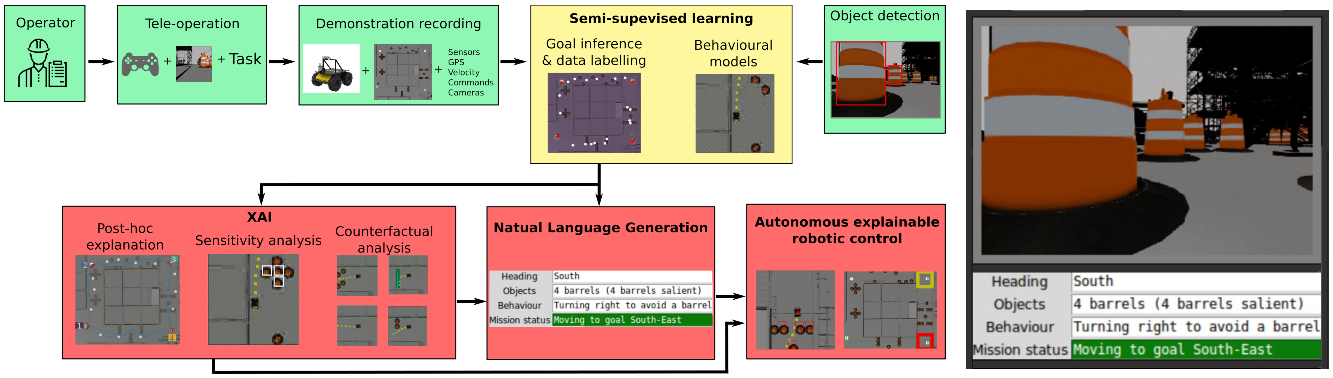
Nevertheless, typically a single robot is not able to perform the multitude of tasks needed for complex operations. Thus, operators often have to supervise several types of robots at the same time (e.g. ground robots with hoses to extinguish fires and aerial drones for inspections). This is exacerbated by the remote nature of the environment, as remotely-controlled robots often instil less trust than those co-located [1, 9]. Maintaining a correct and clear operator mental model is thus necessary for the deployment of multiple adaptive robots simultaneously.

Dynamic robot behaviours may be difficult to understand, in particular to non-experts and novice operators. We provide a way to communicate these clearly through natural language messages synthesised from the semi-supervised learning algorithm. Presenting the robot behaviour in words can increase situation awareness, helping to better understand what the robot is doing and why, as well as helping the operator know when to intercede if the robot is not operating correctly [4, 16].

In this video and paper, we describe a system that combines semi-supervised learning and natural language explanations of behaviour to learn from demonstrations and clarify or verify the behaviours displayed by the robot.

2 BEHAVIOURAL MODEL LEARNING

We train DNNs as behavioural models based on demonstrations carried out by robot operators [19]. In this LfD (learning from demonstration) configuration (Fig. 1a), a user teleoperates a Husky robot in an inspection scenario. The operator receives exteroceptive information about the robot from an on-board camera and drives it controlling linear and angular velocities. We recorded five demonstrations varying the location of obstacles, but maintaining the same points on the remote platform as inspection goals.



(a) System architecture **(b) Explainable inteface**

Figure 1: (a) Explainable learning from demonstration architecture. Green blocks represent input modules to the system. The yellow blocks represent the semi-supervised learning of behavioural models. Red blocks are the main outputs of the system in graphical and natural language form. (b) Interface including Natural Language Generation (NLG).

First, following [3], we extract goals in the demonstrations as the most likely position on the platform to where the robot is being directed. Using a sequential importance sampling algorithm, we extract the goals and parameters of the proportional controllers that satisfy them. We filter the goals using their likelihood of being part of the actual path taken by the robot and their saliency in a path-prediction model. Once the system has inferred the goals, we label each step of the original demonstration to learn goal-based Deep Neural Networks (DNNs). These models take the state of the robot as input and predict the most likely path to be followed by the robot for the next 5 seconds.

We use the behavioural models for different tasks (red blocks in 1a). First, we can control the robot in an MPC (model-predictive control) way. Second, we explain the path prediction from the models with causal inference. Using sensitivity analysis, we are able to point out the objects around the robot that are important for behaviour prediction. Also, using counterfactual analysis, the system can inform a user about the required modifications to follow a specific path [18]. Finally, the combination of these behavioural explanations can be translated to natural language.

3 NATURAL LANGUAGE INTERFACE

Although robots that adapt are more useful in certain environments, this ability to adapt can make their behaviour obscure to non-experts. This decreased understanding can have a negative effect on the operator's trust, and thus on the human-robot collaboration. Improving the robot's transparency is important for explainability and trust [14, 21] and it is crucial for a clear operator mental model, which can prevent issues such as wrong assumptions, misuse, disuse or over-trust [4]. A more faithful mental model also helps to increase the operator's confidence and performance [2, 11].

Previous works have communicated the robot's actions through natural language during and after the tasks [4, 8, 12, 15]. They provide information through automatic reports or real-time dialogue, yet, unlike our case, these robots often have well-defined behaviour that could be initially described by an expert or coded from domain-specific knowledge.

The system, described here (see Fig. 1b) and illustrated in the video generates updates about the actual direction of the robot ('Heading') based on the robot's internal measurement unit (IMU), a list of objects that are sensitive for the prediction ('Objects'), the current behaviour ('Behaviour') based on the predicted path followed by the MPC and the position of salient objects, and an overall mission status. Combining these outputs using template based generation, the system provides easy-to-digest explanations about the robot's behaviour and deviations from its standard trajectory. Some examples are "Turning left to avoid a barrel" or "Making a u-turn towards South-West goal". These are shown on the operator's user interface to help increase understanding and situation awareness.

4 CONCLUSION AND FUTURE WORK

In this demonstration, we have shown a robot that autonomously navigates an environment and produces updates about what it is doing and why. The robot's behaviour is learned from an operator demonstration, from which a program is synthesised and then executed using a semi-supervised learning algorithm. The updates adapt to the robot's behavioural model and can be used to keep operators informed about what the robot is thinking of doing next. In the future, we would like to extend this to more types of updates and allow one to more closely examine the robot's actions through free-chat. Explanations could become crucial if we want to better understand highly autonomous, adaptive robots that show behaviour that may be too complicated even for experts to understand. Evaluating and optimising natural language explanations is also left for future work [5]. Finally, this work could also open the door to live debugging of robots and automatically-synthesised programs, to verify that the robot is acting as it should or explain why it did not correctly learn from the demonstration.

ACKNOWLEDGMENTS

This work was supported by the EPSRC-funded ORCA Hub (EP/R026173/1, 2017-2021) and UKRI Trustworthy Autonomous Systems Nodes in Trust (EP/V026682/1, 2020-2024) and in Governance and Regulation (EP/V026607/1, 2020-2024).

REFERENCES

[1] Wilma A. Bainbridge, Justin Hart, Elizabeth S. Kim, and Brian Scassellati. 2008. The Effect of Presence on Human-Robot Interaction, In 17th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN). *Proceedings IEEE Int. Symp. on Robot and Human Interactive Communication*, 701–706. https://doi.org/10.1109/ROMAN.2008.4600749

[2] Pierre Le Bras, David A. Robb, Thomas S. Methven, Stefano Padilla, and Mike J. Chantler. 2018. Improving User Confidence in Concept Maps: Exploring Data Driven Explanations. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*. ACM, 1–13. https://doi.org/10.1145/3173574.3173978

[3] Michael Burke, Svetlin Penkov, and Subramanian Ramamoorthy. 2019. From Explanation to Synthesis: Compositional Program Induction for Learning From Demonstration. *Robotics: Science and Systems (R:SS)* (June 2019). https://doi.org/10.15607/RSS.2019.XV.015

[4] Francisco J. Chiyah Garcia, David A. Robb, Atanas Laskov, Xingkun Liu, Pedro Patron, and Helen Hastie. 2018. Explainable Autonomy: A Study of Explanation Styles for Building Clear Mental Models. In *Proceedings of The 11th International Natural Language Generation Conference (INLG'18)*. ACM, Tilburg, The Netherlands, 99–108. http://www.aclweb.org/anthology/W18-65#page=119

[5] Miruna Clinciu, Arash Eshghi, and Helen Hastie. 2021. A Study of Automatic Metrics for the Evaluation of Natural Language Explanations. In *Proceedings of the European Association of Computational Linguistics (EACL'21)*.

[6] Helen Hastie, Katrin Lohan, Mike Chantler, David A Robb, Subramanian Ramamoorthy, Ron Petrick, Sethu Vijayakumar, and David Lane. 2018. The ORCA Hub: Explainable Offshore Robotics through Intelligent Interfaces. In *Proceedings of Explainable Robotic Systems Workshop, ACM HRI Conference*. 1–2.

[7] Young-Sik Kwon and Byung-Ju Yi. 2012. Design and motion planning of a two-module collaborative indoor pipeline inspection robot. *IEEE Transactions on Robotics* 28, 3 (June 2012), 681–696. https://doi.org/10.1109/TRO.2012.2183049

[8] Oliver Lemon, Anne Bracy, Alexander Gruenstein, and Stanley Peters. 2001. The WITAS multi-modal dialogue system. In *Seventh European Conference on Speech Communication and Technology*. 1559–1562.

[9] Jamy Li. 2015. The benefit of being physically present: A survey of experimental works comparing copresent robots, telepresent robots and virtual agents. *International Journal of Human-Computer Studies* 77 (2015), 23–37. https://doi.org/10.1016/j.ijhcs.2015.01.001

[10] Jinke Li, Xinyu Wu, Tiantian Xu, Huiwen Guo, Jianquan Sun, and Qingshi Gao. 2017. A novel inspection robot for nuclear station steam generator secondary side with self-localization. *Robotics and Biomimetics* 4, 1 (2017), 26. https://doi.org/10.1186/s40638-017-0078-y

[11] Brian Y. Lim, Anind K. Dey, and Daniel Avrahami. 2009. Why and why not explanations improve the intelligibility of context-aware intelligent systems. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Boston, MA, USA) *(CHI '09)*. 2119–2129. https://doi.org/10.1145/1518701.1519023

[12] Teruhisa Misu, Antoine Raux, Ian Lane, Joan Devassy, and Rakesh Gupta. 2013. Situated multi-modal dialog system in vehicles. In *Proceedings of the 6th workshop on Eye gaze in intelligent human machine interaction: gaze in multimodal interaction - GazeIn '13*. ACM Press, New York, New York, USA, 25–28. https://doi.org/10.1145/2535948.2535951

[13] Keiji Nagatani, Seiga Kiribayashi, Yoshito Okada, Kazuki Otake, Kazuya Yoshida, Satoshi Tadokoro, Takeshi Nishimura, Tomoaki Yoshida, Eiji Koyanagi, and Mineo Fukushima. 2013. Emergency response to the nuclear accident at the Fukushima Daiichi Nuclear Power Plants using mobile rescue robots. *Journal of Field Robotics* 30, 1 (2013), 44–63. https://doi.org/10.1002/rob.21439

[14] Birthe Nesset, David A. Robb, José Lopes, and Helen Hastie. 2021. Transparency in HRI: Trust and Decision Making in the Face of Robot Errors. In *Proceedings of ACM/IEEE International Conference on Human-Robot Interaction (HRI'21)*.

[15] Vittorio Perera, Sai P. Selveraj, Stephanie Rosenthal, and Manuela Veloso. 2016. Dynamic generation and refinement of robot verbalization. In *25th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN)*. IEEE, New York, NY, USA, 212–218. https://doi.org/10.1109/ROMAN.2016.7745133

[16] David A. Robb, Francisco J. Chiyah Garcia, Atanas Laskov, Xingkun Liu, Pedro Patron, and Helen Hastie. 2018. Keep Me in the Loop: Increasing Operator Situation Awareness through a Conversational Multimodal Interface. In *Proceedings of the 20th ACM International Conference on Multimodal Interaction (ICMI'18)*. ACM, Boulder, Colorado, USA, 384–392. https://doi.org/10.1145/3242969.3242974

[17] Amit Shukla and Hamad Karki. 2016. Application of robotics in onshore oil and gas industry—-A review part I. *Robotics and Autonomous Systems* 75 (2016), 490–507. https://doi.org/10.1016/j.robot.2015.09.012

[18] Simón C. Smith and Subramanian Ramamoorthy. 2020. Counterfactual Explanation and Causal Inference in Service of Robustness in Robot Control. arXiv:2009.08856 [cs.RO]

[19] Simón C. Smith and Subramanian Ramamoorthy. 2020. Semi-supervised Learning From Demonstration Through Program Synthesis: An Inspection Robot Case Study. *Electronic Proceedings in Theoretical Computer Science* 319 (jul 2020), 81–101. https://doi.org/10.4204/EPTCS.319.7

[20] Cuebong Wong, Erfu Yang, Xiu-Tian T. Yan, and Dongbing Gu. 2017. An overview of robotics and autonomous systems for harsh environments. In *2017 23rd International Conference on Automation and Computing (ICAC)*. IEEE, Huddersfield, UK, 1–6. https://doi.org/10.23919/IConAC.2017.8082020

[21] Robert H. Wortham, Andreas Theodorou, and Joanna J. Bryson. 2017. Robot transparency: Improving understanding of intelligent behaviour for designers and users. In *Towards Autonomous Robotic Systems: 18th Annual Conference, TAROS 2017 (Lecture Notes in Artificial Intelligence)*, Yang Gao, Saber Fallah, Yaochu Jin, and Constantina Lakakou (Eds.). Springer, Guildford, UK, 274–289. https://doi.org/10.1007/978-3-319-64107-2_22