ANITI
ARTIFICIAL & NATURAL INTELLIGENCE TOULOUSE INSTITUTE

Contact: Caroline Chanel
and Nicolas Drougard
firstname.lastname@isae.fr

ISAE SUPAERO
Institut Supérieur de l'Aéronautique et de l'Espace

Université Fédérale
Toulouse Midi-Pyrénées

PhD proposition: Automated POMDP model learning for driving Human-Robot Interaction

Introduction

Mixed-initiative missions involving Human-Robot Interactions (HRI) have become increasingly common (Adams et al., 2004). As a result, they play a significant role in the autonomous systems research and applications (autonomous UAV, autonomous cars, avionics, etc). The idea of this thesis proposition is to compute strategies for driving human-robot systems by using crowdsourcing data. For instance, in the mixed-initiative framework (Allen et al., 1999; Adams et al., 2004), the role of each agent (human or artificial) is defined according to their recognized skills. In our point of view, it should be preferable to define the different roles according to their recognized skills and their current capabilities (Gateau et al., 2016).

In this context, it is necessary to monitor the agents' performance during interaction, and, for instance, decide which task should be performed by each agent. To be able to predict long-term human-system performance one could rely on an HRI model, *i.e.* a description of the human-system interaction dynamics (Drougard et al., 2017; Charles et al., 2018). The goal of this thesis is to develop a framework for building an accurate and efficient model from data in order to optimize a Human-Robot Interaction strategy.

Content

This PhD topic addresses the issue of learning and planning under uncertainty. Indeed an HRI model may rely on probability functions depending on state variables of interest. One of the important issue is to automatically select the most promising state variables (Castelletti et al., 2011), *i.e.* the ones that influence the utility of performing a supervision action. Another issue is to automatically define a good granularity (Li et al., 2006) of the selected variable spaces in order to be able to learn precise enough probability values defining the HRI dynamics described in the form of a tractable sequential decision-making problem.

Moreover, it is worth to say that the human operator behavior is possibly non-deterministic, however her/his (cognitive) state is also partially observable by definition. Therefore monitoring the human operator's (cognitive) state and the related performance is a challenging task. Another issue in this thesis proposition is to include hidden state variables in the model, in order to describe the human operator (Gateau et al., 2016; de Souza et al., 2015). Such hidden state variables should then be estimated during HRI from observable data (de Souza et al., 2015; Drougard, 2015).

In a nutshell, this thesis aims to propose a general framework to learn a Partial Observable Markov Decision Process (POMDP, (Cassandra, 1998)) model from action-observation sequences, and to solve such a model in order to obtain a reliable strategy (Silver and Veness, 2010; Kurniawati et al., 2008). The strategy, also called policy, will then be explored to drive mixed-initiative human-robot interaction.

Existing proof-of-concept scenario

In order to set up a mixed-initiative mission involving HRI, the existing proof of concept scenario rely on a computer-based simulation environment called *Firefighter Robot Game* (Drougard et al., 2017) available at the following address: `http://robot-isae.isae.fr`. This mission has been designed to collect human (behavioral and physiological) data and to test

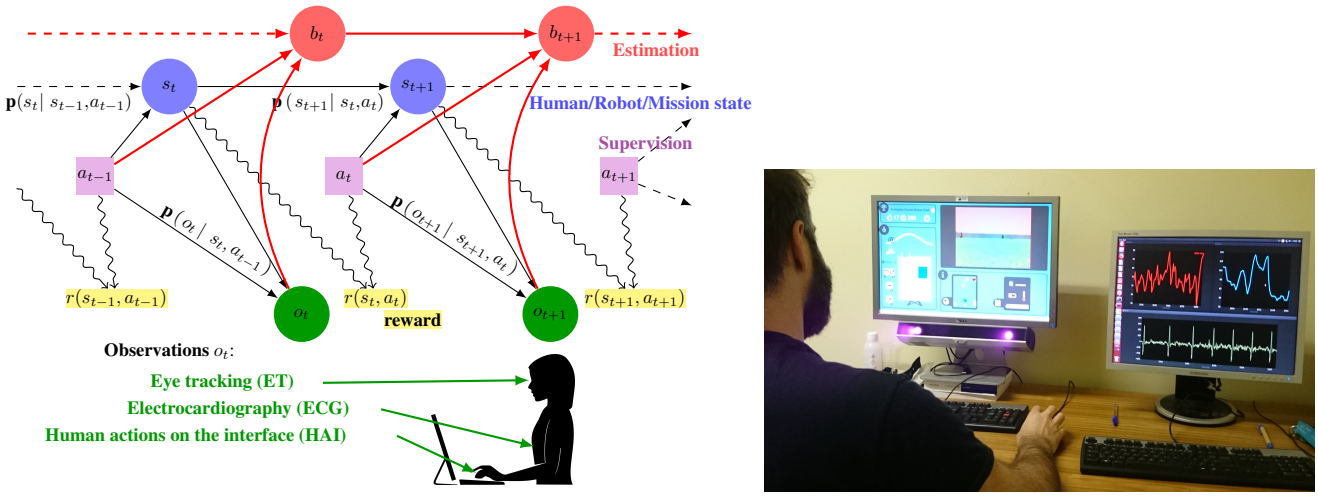

Figure 1: POMDP model schema relying on physiological markers and human actions on the interface.

supervision strategies aiming at driving the Human-Robot system by defining robot and human tasks, or launching alarms, and thus adapt the system to the situation and the human behavior. Such strategies can be computed by planning under uncertainty algorithms maximizing the entire system's performance. Ultimately the produced strategies could be tested with a real human-robot team and environment in our laboratory.

In this scenario, the operator is immersed in a stressful situation in which burning trees must be extinguished with a remote controlled firefighter robot. At the same time the operator must ensure that an external water tank is full enough to supply the robot. This last task is not easy to perform correctly since leaks appear on the tank and the filling tap is unstable. The robot can go into autonomous mode and thus follow a fixed policy depending on the environment and its own state. Its temperature increases when it is close to flames and its battery must be recharged in time since its level decreases during the mission. The water tank on the robot is also not unlimited and must therefore be recharged using the external water tank. All these constraints can degrade the mental state of the human operator, in particular because overheated temperature and low battery prematurely ends the mission (failure). The goal of the mission is to extinguish as many fires as possible in a limited time. The robot, when in manual mode, is controlled by the arrows (resp. the space bar) of the keyboard for the movements (resp. for the jets of water). The external water tank level can be managed with several letter keys or with clicks on the buttons of the graphical user interface. In this environment the appearance of fires can be experienced as a danger. The limited mission time, the temperature, the battery as well as the score can be a source of stress and pressure. As a result, the entire Robot Firefighter task is hard enough to generate variations on the engagement of the human operator and possible undesirables mental states.

Previous work (Charles et al., 2018) have proposed a computation pipeline to approach a Markov Decision Process (MDP) model of this proof-of-concept scenario. The MDP model was approximated based on among 80 hours of game play. The model was evaluated in simulation and the achieved results are promising. However, the proposed approach have some limitations. In parallel, another previous work (Chanel, 2019) investigated the predictive power of physiological markers useful to estimate the human operator engagement during HRI. The 18 participants have performed 4 times the Firefighter Robot mission in our lab facilities equipped with an Electrocardiogram (ECG) and an Eye-tracker (ET). First results are promising in the sense that physiological markers can be exploited to produce informative observations (about the human operator) that would most certainly be useful in a POMDP model.

These previous works pave the way towards systems that are able to take into account the human operator in order to improve performance in human-robot interactions. However the previous acquired data (crowdsourcing and lab experiments) have unbalanced sizes. Another challenge for future work would be to adequately merge these data sets to learn the parameters of the POMDP model.

PhD candidate's profile:

- Mathematics, Probability-Statistics, Machine Learning, Reinforcement Learning, Artificial Intelligence background;
- Strong programming skills;
- Autonomous, hard-working, problem-solver;
- Interested in Human-Machine Interaction.

For information, the main libraries useful for this PhD: Scikit-learn – https://scikit-learn.org, PROST – https://bitbucket.org/tkeller/prost, Pytorch – https://pytorch.org, Pomegranate – https://pomegranate.readthedocs.io, RDDLSim – https://github.com/ssanner/rddlsim, APPL – https://bigbird.comp.nus.edu.sg/pmwiki/farm/appl/

Additional Information

- Salary: This PhD thesis is financially supported by the ANITI Institute which offers a competitive net salary of 2096 euros per month with some teaching (64 hours per year on average) .
- Starting date: January.
- Duration: 36 months
- Supervisors: Dr Caroline P.C. Chanel and Dr Nicolas Drougard, ISAE-SUPAERO, Université de Toulouse, France.
- Collaborators: ANITI Chair holder Professor Frédéric Dehais and Dr Raphaëlle N. Roy.
- Application procedure: Formal applications should include a detailed CV, a motivation letter, at least one reference letter, and transcripts of degrees. Samples of published research by the candidate will be a plus.

References

Adams, J. A., Rani, P., and Sarkar, N. (2004). Mixed initiative interaction and robotic systems. In *AAAI Workshop on Supervisory Control of Learning and Adaptive Systems*, pages 6–13.

Allen, J., Guinn, C. I., and Horvtz, E. (1999). Mixed-initiative interaction. *IEEE Intelligent Systems and their Applications*, 14(5):14–23.

Cassandra, A. R. (1998). A survey of pomdp applications. In *Working notes of AAAI 1998 fall symposium on planning with partially observable Markov decision processes*, volume 1724.

Castelletti, A., Galelli, S., Restelli, M., and Soncini-Sessa, R. (2011). Tree-based variable selection for dimensionality reduction of large-scale control systems. In *2011 IEEE Symposium on Adaptive Dynamic Programming and Reinforcement Learning (ADPRL)*, pages 62–69. IEEE.

Chanel, C.P.C., R. R. D. F. . D. N. (2019). Towards mixed-initiative human-robot interaction: Assessment of discriminative physiological and behavioral features for performance prediction. *Frontiers in Robotics & AI (under revision)*.

Charles, J.-A., Chanel, C. P., Chauffaut, C., Chauvin, P., and Drougard, N. (2018). Human-agent interaction model learning based on crowdsourcing. In *Proceedings of the 6th International Conference on Human-Agent Interaction*, pages 20–28. ACM.

de Souza, P. E. U., Chanel, C. P. C., and Dehais, F. (2015). Momdp-based target search mission taking into account the human operator's cognitive state. In *2015 IEEE 27th International Conference on Tools with Artificial Intelligence (ICTAI)*, pages 729–736. IEEE.

Drougard, N. (2015). *Exploiting imprecise information sources in sequential decision making problems under uncertainty*. PhD thesis, Université de Toulouse, ISAE-SUPAERO.

Drougard, N., Ponzoni Carvalho Chanel, C., Roy, R. N., and Dehais, F. (2017). Mixed-initiative mission planning considering human operator state estimation based on physiological sensors. In *IROS-2017 workshop on Human-Robot Interaction in Collaborative Manufacturing Environments (HRI-CME)*.

Gateau, T., Chanel, C. P. C., Le, M.-H., and Dehais, F. (2016). Considering human's non-deterministic behavior and his availability state when designing a collaborative human-robots system. In *2016 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pages 4391–4397. IEEE.

Kurniawati, H., Hsu, D., and Lee, W. S. (2008). Sarsop: Efficient point-based pomdp planning by approximating optimally reachable belief spaces. In *Robotics: Science and systems*, volume 2008. Zurich, Switzerland.

Li, L., Walsh, T. J., and Littman, M. L. (2006). Towards a unified theory of state abstraction for mdps. In *ISAIM*.

Silver, D. and Veness, J. (2010). Monte-carlo planning in large pomdps. In *Advances in neural information processing systems*, pages 2164–2172.