



Robot Learning from Unlabelled Teaching Signals

Jonathan Grizou, Manuel Lopes, Pierre-Yves Oudeyer

► **To cite this version:**

Jonathan Grizou, Manuel Lopes, Pierre-Yves Oudeyer. Robot Learning from Unlabelled Teaching Signals. HRI 2014 Pioneers Workshop, Mar 2014, Bielefeld, Germany. hal-00963725

HAL Id: hal-00963725

<https://hal.inria.fr/hal-00963725>

Submitted on 24 Mar 2014

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Robot Learning from Unlabelled Teaching Signals

Jonathan Grizou
Flowers Team
INRIA / ENSTA-Paristech
France
jonathan.grizou@inria.fr

Manuel Lopes
Flowers Team
INRIA / ENSTA-Paristech
France
manuel.lopes@inria.fr

Pierre-Yves Oudeyer
Flowers Team
INRIA / ENSTA-Paristech
France
pierre-yves.oudeyer@inria.fr

1. INTRODUCTION

Human Robot Interaction (HRI) is the multidisciplinary study of interactions between humans and robots. Researchers in the field investigate how robots properties — such as embodiment, speech, gesture, facial expression, personality — can change the behaviour, feelings or attitudes of humans; how robots can produce and recognise communicative acts and how robots and humans can learn from each other [2, 3]; among others.

Among them (a) perception and (b) task learning are vastly investigated areas which rely mainly on the machine learning and artificial intelligence community. While impressive results have been obtained in each domain [1, 10], most work considered them separately making important assumptions:

- (a) In perceptual learning research applied to HRI, human’s communicative signals are learnt to be associated to their respective meanings. Such learning often rely on supervised classification or regression methods, which implies collecting signals samples associated with their respective meanings. This latter requirement implies **the robot is aware of the communicative goal of the human**.
- (b) In robot learning from human interaction most work has focused on how to extract statistical task models from human teaching signals [8]. Therefore a usual assumption is that **the robot understands the meanings of human’s communicative signals**.

Their respective assumptions makes those two lines of research incompatible. On the one hand, working on perceptual learning, i.e. learning the signal-to-meaning mapping, requires the robot to know the task. On the other hand, teaching a new task to a robot requires the robot to already know the signal-to-meaning mapping. Consequently it is impossible for a user to interact — from scratch — with a robot using his own preferred teaching signals, the user must comply to the use of pre-defined ones. This paper describes a preliminary approach allowing a robot to be instructed a new unknown task by a human using communicative signals initially totally unknown to the robot. In other words, we address the problem of removing the need for calibration.

The question of how a robot can learn to interpret personalised and potentially noisy teaching signals has not been

much explored. In a preliminary work [9], we presented a computational approach addressing this problem by considering a finite space of teaching signals in simulation but requiring to bootstrap the system with known symbols. Later [7], we released the need for bootstrapping and allow the teacher to use any kind of signal that can be represented as a fixed length feature vector, which is better suited for HRI scenarios.

We provide an intuition on how the algorithm works and report preliminary results.

2. INTUITION

For the present discussion, we restrain our analysis to scenarios where the user does not actively deliver commands to the robot, but only delivers feedback about actions performed by the robot. Such feedback is given by an unlabelled communicative signals than can be either symbolic, e.g. button presses, or represented as a fixed length feature vector, e.g. spoken words. Signals are **not** converted to a symbolic meaning. The robot needs to actively execute actions to solve the tasks which implies learning the signal-to-meaning association.

This control can be exemplified for a reaching task, where the user wants the robot to reach a target position unknown at start. The robot performs several actions (e.g. moving left or right), and receives unlabelled feedback signals from the user. The feedback signals are generated as a response to the execution of an action in a state according to the true unknown task the user wants to solve. For instance, binary feedback signals simply encode whether the action executed is correct or incorrect according to the user intended task. The key point is that these signals are generated from an underlying model that for binary signals has two different classes.

To solve this problem, we must exploit an other source of information, namely task constraints. Task constraints are properties of the environment that limit the space of possible task, e.g. of possible target positions. This set of hypothetical tasks enable us to create a set of hypothetical signal-to-meaning models. Since the right task will assign the right labels to the signals, while the other tasks will gradually mix them, estimating models’ likelihood is a good measure to identify the user’s intended task.

3. PRELIMINARY RESULTS

We construct a small size pick-and-place task with a real robot that has 4 actions available: *rotate left*, *rotate right*,

grasp cube and *release cube*. This robot is going to be programmed using natural speech with words a priori unknown to the robot. The teacher is facing the robot and chooses a specific arrangement of cube, i.e. a specific task, it wants the robot to build. It then decides one word to use as positive feedback and one as negative feedback and starts teaching the robot. Once the robot has understood the first task, using the method presented above, it has also understood the signal-to-meaning model. We can freeze the corresponding classifier and start learning a new task.

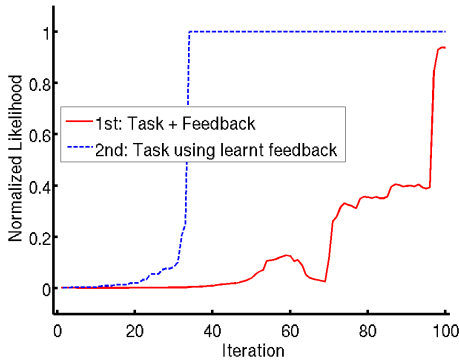


Figure 1: Evolution of the probability of the taught task. 1) The robot learns a task from unlabelled speech feedback. 2) By freezing the classifier corresponding to the first task, the user teaches the robot a new task faster.

Fig. 1 shows results from this setting. In the first run it takes about 100 iterations for the robot to learn both the task and the signal-to-meaning mapping. Whereas in the second run, when reusing knowledge from the first one, the robot is able to learn a new task faster, in about 30 iterations, meaning that it has well found the two clusters in our feature space as well as the mapping to their corresponding meanings.

4. LIMITATIONS AND PERSPECTIVES

We presented a new approach for HRI which enables a user to use his own preferred teaching signals when interacting with a robot, e.g. using his own language or defining his own finger movements. The claim that it would be more intuitive or natural to interact with a robot this way is far from being obvious and should be tested with care. When interacting with machine, including robots, we are often, if not always, told how to use them. Do people want to have an open-ended choice about how to interact with machines? Would they be more efficient? Investigating such question is part of our future work.

Our method remove the necessity of a calibration phase. Among the many application requiring calibration, Brain Computer Interaction (BCI) is one of the most challenging. We are currently investigating in this direction [6]. BCI has the advantage to make obvious the fact that we cannot ask the user to comply to the use of pre-defined signals.

An important assumption of our method is that it is possible to define a finite — and reasonable — set of task hypothesis. This assumption is limiting for many theoretical problems but many useful real word application can still benefit from it. An eligible scenario is the problem of grasping, on a table,

one object among a finite set of objects. In this scenario the set of hypothesis consist of all the objects on the table.

While this is not the main target, the work presented in this paper is also relevant with regards to infant social development and learning, as well as in adult mutual adaptation of social cues. This has been the subject of experiments in experimental semiotics [4], such as in the work of Griffiths et al. [5] who conducted an experiment with human learners learning the meaning of unknown symbolic teaching signals. An innovative direction would be to embed the algorithmic principles introduced in this paper for experimental linguistics studies.

Acknowledgement

Work (partially) supported by INRIA, Conseil Régional d’Aquitaine and the ERC grant EXPLORERS 24007.

5. REFERENCES

- [1] B. D. Argall, S. Chernova, M. Veloso, and B. Browning. A survey of robot learning from demonstration. *Robotics and Autonomous Systems*, 57(5):469–483, 2009.
- [2] K. Dautenhahn. Socially intelligent robots: dimensions of human–robot interaction. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 362(1480):679–704, 2007.
- [3] T. Fong, I. Nourbakhsh, and K. Dautenhahn. A survey of socially interactive robots. *Robotics and autonomous systems*, 42(3):143–166, 2003.
- [4] B. Galantucci. Experimental semiotics: a new approach for studying communication as a form of joint action. *Topics in Cognitive Science*, 1(2):393–410, 2009.
- [5] S. Griffiths, S. Nolfi, G. Morlino, L. Schillingmann, S. Kuehnel, K. Rohlfing, and B. Wrede. Bottom-up learning of feedback in a categorization task. In *Development and Learning and Epigenetic Robotics (ICDL), 2012 IEEE International Conference on*, pages 1–6. IEEE, 2012.
- [6] J. Grizou, I. Iturrate, L. Montesano, M. Lopes, and P.-Y. Oudeyer. Zero-calibration bmis for sequential tasks using error-related potentials. In *IROS 2013 Workshop on Neuroscience and Robotics*, Tokyo, Japon, 2013.
- [7] J. Grizou, M. Lopes, and P.-Y. Oudeyer. Robot Learning Simultaneously a Task and How to Interpret Human Instructions. In *Joint IEEE International Conference on Development and Learning and on Epigenetic Robotics (ICDL-EpiRob)*, Osaka, Japan, 2013.
- [8] F. Kaplan, P.-Y. Oudeyer, E. Kubinyi, and A. Miklósi. Robotic clicker training. *Robotics and Autonomous Systems*, 38(3):197–206, 2002.
- [9] M. Lopes, T. Cederbourg, and P.-Y. Oudeyer. Simultaneous acquisition of task and feedback models. In *Development and Learning (ICDL), 2011 IEEE International Conference on*, volume 2, pages 1–7. IEEE, 2011.
- [10] Z. Zeng, M. Pantic, G. I. Roisman, and T. S. Huang. A survey of affect recognition methods: Audio, visual, and spontaneous expressions. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 31(1):39–58, 2009.