AI for HRI: Learning From Human Preferences And Intentions For Smooth And Adaptive Handover

Francesco Iori, Gojko Perovic, Matteo Conti, Francesca Cini, Angela Mazzeo, Marco Controzzi, Egidio Falotico
Sant'Anna School of Advanced Studies, The Biorobotics Institute
f.iori@santannapisa.it, g.perovic@santannapisa.it

Abstract

Human-Robot Interaction research should consider possible applications to real-world problems, and thus take into account the unanticipated factors that might hinder the interaction. In a task of robot-to-human handover, a scenario where human movements can be unexpectedly perturbed or interrupted is considered. An approach derived from Dynamic Movement Primitives is proposed for online trajectory generation of a robotic arm. Two Machine Learning modules are presented to endow the robot controller with adaptive capabilities. First, a Bayesian Optimization-based preferential learning algorithm is used to tune the robot behavior on human feedback. Then, a short-term predictor is employed to facilitate anticipatory control. By predicting human-hand displacement and anticipating the trajectory of the handover, we demonstrate that the proposed method can be robust and safe for online trajectory generation in handover tasks.

1 Introduction

While there is a growing interest in Human-Robot Interaction (HRI) research, practical implementations, especially for industrial use, remain limited. Most commonly, in industry, robots are contained in isolated workspaces, separated from human workers, due to safety and efficiency constraints. However, to obtain human-like collaboration capabilities, the development of robots that occupy shared workspaces and participate in joint tasks is compulsory. Although there is extensive research on HRI, laboratory conditions tend to be more structured, to focus on certain aspects of the interaction, as opposed to practical interactions which may include unexpected events. Thus, it might be necessary to move towards paradigms able to work in less structured settings and to consider scenarios that more closely match practical applications to develop pragmatic HRI implementations.

A typical problem in HRI is executing a handover: passing an object between two agents. As this simple action can be considered fundamental for most conceivable physical collaborations, it has received broad attention from the research community. However, a handover requires at one time many different problems: partner and objects recognition, interpretation of partners intention, grasp selection and control, trajectory planning, and more. Thus, the reliable execution of a handover with the level of smoothness akin to humans still represents an open problem [Ortenzi *et al.*, 2021].

To address some of the aforementioned issues we turned to a combination of dynamical systems and Machine Learning (ML). Dynamical systems present a solid baseline for online trajectory generation. In this way, the trajectory is not completely planned beforehand, but it is generated in real-time as the robot executes the motion. This allows for reaction to what happens in the environment on the fly, without the need to re-plan the motion. On the other hand, ML is employed to address human-related aspects of interaction, which would otherwise be difficult to model.

In this work, we focus on the problem of generating a suitable trajectory for the robot to pass the object, while being able to react to unpredictable perturbations that may happen and adapt to the user's preference. The methods proposed are focused on reacting given the *permanence of the intention to perform the handover* from the robot. An example of such handover is illustrated in Figure 1. As an example, consider a high-level controller that, interpreting the action of the human partner, decides when to initiate or to stop the execution of the handover. Such controller could fail or be too slow to recognize when the handover should be interrupted or adjusted, or can start the handover at the wrong time (e.g. too soon). The methods presented here can be integrated with such a high-level controller to lower its performance requirements, while still keeping a smooth and safe behaviour.

2 Basic framework

2.1 Trajectory generation with dynamical systems

The methods applied in this work are a generalization of Dynamic Movement Primitives (DMP), introduced by [Ijspeert *et al.*, 2002]. The trajectory is considered as generated by the evolution of a dynamical system. As an example, we can consider a one-dimensional trajectory $x(t)$ generated by a second-order dynamical system similar to a mass-spring-damper system:

$$\tau^2 \ddot{x} = \alpha_x(\beta_x(g - x) - \tau \dot{x}) + f_{ext} \quad (1)$$

with f_{ext} an external forcing term. If we consider that f_{ext} vanishes over time (typically obtained by introducing an ad-

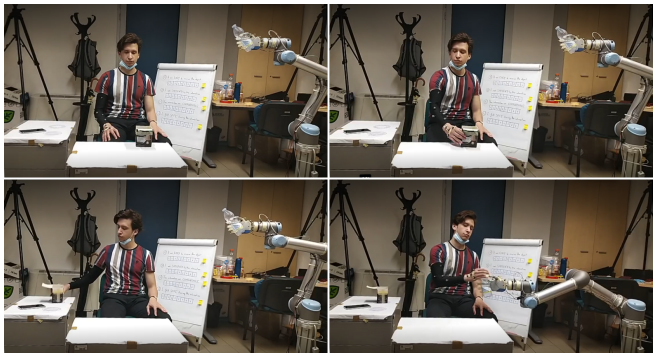


Figure 1: Example of robust handover. The robot starts the execution in the first frame. Then, human performs a secondary task of taking an object close to handover location and placing it to the side. Finally, the human reaches for the handover object.

ditional variable, called the *phase* of the system), for positive values of α_x and β_x, the state will smoothly converge to the goal g. The forcing term can also be learned, to replicate a desired trajectory before convergence to the goal. Furthermore, additional terms can be added to modify the behaviour of the system depending on external cues. In this work we consider a more generalized version of the system above:

$$\dot{\boldsymbol{y}} = f(\boldsymbol{y}, \boldsymbol{\theta}, \boldsymbol{u}) \tag{2}$$

with a state $\boldsymbol{y} = [x, \dot{x}, ...]^T$, external input u, and θ the vector of parameters of the system.

2.2 Preferential learning

Due to the inherent flaws in human evaluations, learning from human input is challenging. Absolute human feedback is usually noisy and unreliable, suffering from drift (scale shifting over time) and anchoring (early interactions are deemed as more important) [Chu e Ghahramani, 2005; Brochu *et al.*, 2010]. Furthermore, internal scales between individuals could be vastly different. To overcome these issues, relative feedback can be employed. Thus, human partners can be asked how did the most recent interaction compare to the previous one. Given the preferences, the *probit* model can be applied, that fits relative, binary, feedback to utility function u[Chu e Ghahramani, 2005; Brochu *et al.*, 2010]. Let's consider a data set of ranked pairs:

$$\mathcal{D} = \{r_i \succ c_i; i = 1, ..., m\} \tag{3}$$

where $r_i, c_i \in \Theta$ are points in the parameter space. After collecting the data, a zero-mean non-parametric Gaussian Process (GP) prior can be fitted as:

$$\mathcal{P}(\mathbf{u}) = |2\pi\mathbf{K}|^{\frac{1}{2}} \exp\left(\frac{1}{2}\mathbf{u}^T\mathbf{K}^{-1}\mathbf{u}\right) \tag{4}$$

where $\mathbf{u} = |u(\theta_1), u(\theta_2), ..., u(\theta_n)|^T$ and $\mathbf{K}$ is the $n \times n$ covariance matrix (n is the number of instances) [Chu e Ghahramani, 2005]. To estimate the posterior distribution of u, model is fit:

$$\mathcal{P}(\mathbf{u}|\mathcal{D}) \propto \mathcal{P}(\mathbf{u}) \prod_{i=1}^{m} \mathcal{P}(r_i > c_i | u(r_i), u(c_i)) \tag{5}$$

This problem can be viewed as an optimization of an expensive-to-estimate black-box function, and Bayesian Optimization (BO) can be effectively employed [Brochu *et al.*, 2010].

3 Adapting to user preferences with preference learning

The first method conceived to react to perturbations during a handover was based on the estimated distance between the partner's hand and the final handover position (where the hands should meet). Two coupling terms were added to the dynamical system, that slowed down the evolution of the trajectory depending on the distance and its time derivative, with a total of 4 tunable parameters. The idea behind the proposed coupling terms is that the dynamics of the distance from the final handover position can be useful to discriminate the intention to hand over an object. On the other side, the terms can slow down the robot if the human hand is not approaching for the handover. However, there was no easy or intuitive correlation between the value of these parameters and the robot's behaviour.

The task devised to test the controller was composed to represent a possible worst-case scenario: it required the participant to reach for a position close to the (known) final handover position, before going for the actual handover after the execution of a different sub-task. As during all these subtasks the robot was active in handover mode (simulating a high-level controller that activated the handover too soon), if not tuned correctly the controller could be tricked by the first motion to perform the handover at full speed. An example of the task is shown in Figure 1.

In this framework, preference-based Bayesian optimization has been applied to tune the behaviour of the controller, given by its four parameters, directly by the human user's feedback in a perturbed handover scenario. While the preferential feedback addresses the inherent variance introduced by noise in human evaluation, preferences do not convey much information. In virtual environments, where these types of methods are usually applied, this can be addressed by providing users with galleries (a set of multiple examples) to choose from [Brochu *et al.*, 2010]. However, real-world interactions are constrained by sequential experiences and limited by time and memory constraints of the users. We proposed a relative scale with a periodic refresh to alleviate these drawbacks. The scale represents a seven-point scale that ranges from "Strongly Worse" to "Strongly Better" in terms of relative preferences. Interactions are carried out in q sampled point batches. As a result, instead of a two-by-two approach, comparisons are made between the q number of points, increasing the amount of information acquired from each interaction. The scale is refreshed after each batch, eliminating the user's previous feedback. Following that, the user is shown the previous best-observed point $\theta*$, and a fresh set of sampled interactions begins. The benefits of the refresh are twofold: first, as the scale does not carry absolute values, impediments of drift and anchoring are removed; second, the burden on participants' memory is minimized, allowing them to focus on the relationship between the few most recent in-

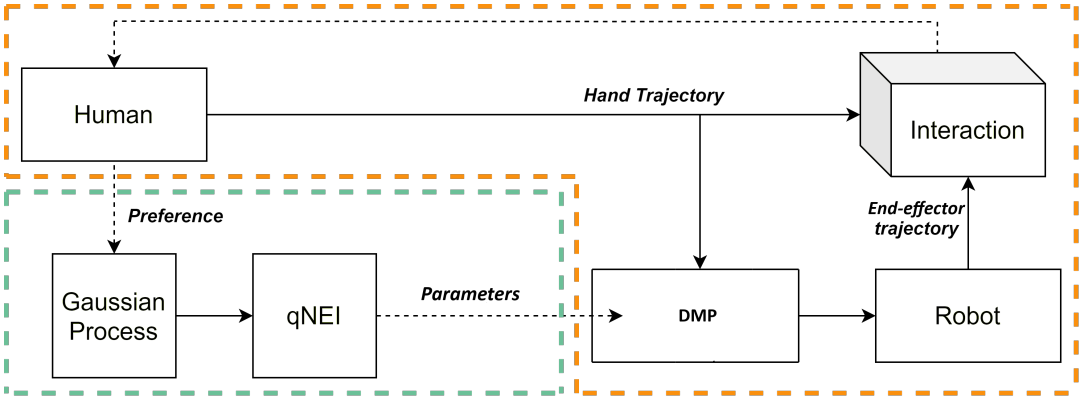


Figure 2: Block diagram of the preference learning loop. Red loop is online trajectory generation. Blue loop is preference learning that happens after each interaction.

teractions. In this way, we aim to increase the impact of each interaction, which in a practical scenario are often very limited. A schematic of the proposed controller and learning loop is shown in Figure 2.

Nine participants participated in this study and performed 15 interactions each. After these interactions participants answered the questionnaire:

- Were you able to find a satisfactory combination of parameters? (Yes: 6; Maybe: 3; No: 0)
- Did you feel that the robot was improving? (Yes: 5; Maybe: 1; No: 3)

Thus, giving encouraging results to more formally examine the application of BO-based algorithms for preferential learning in robotic applications.

While the preference in reactive behavior varied in the overall speed and timing, participants seemed to converge to reactive controllers, as opposed to non-reactive, straightforward, trajectories. This is likely due to the fact that in the proposed scenario cooperative effort is appreciated as it more accurately mimics human behavior, leading to smoother interactions.

4 Short term predictions for robust handover

The main shortcoming of the trajectory generation method introduced in the previous section was the dependency on an estimate of the goal. To extend the robust behaviour to a more general setting and avoid relying extensively on an accurate goal estimate, a novel method inspired by the concept of anticipatory control has been devised. The idea is to endow the robot with "knowledge" of what should happen during a handover and use any mismatch with what really happens to react. Again, the base trajectory is generated with a dynamical system to produce minimum-jerk reaching motion.

To make the robot able to distinguish whether the human is trying to perform a handover or not, two predictors were trained:

1. a short-term predictor of the future displacement of the human hand;
2. a predictor of the direction in which the human hand should move if it were to go for a handover.

The discrepancy between these two measures can then be exploited to detect the intent of a handover (or lack of), as shown in Figure 4. This approach removed the dependency on the final position of the handover, effectively decoupling the two problems of a) where should the agent reach, and b) when should the agent slow down or react. To also address the first question, based on the work of [Prada *et al.*, 2014], the robot was made to converge to the predicted position of the human hand. A diagram of the architecture is shown in Figure 3.

As the two predictors depend only on the past positions of the human hand and the end-effector of the robot, the method is simple, robust, and independent from the specific setting or acquisition system. Furthermore, the two predictors were implemented and tested to work at 30Hz, a rate typical of machine-learning camera-based vision systems (e.g. for hand tracking, gesture detection, skeleton pose reconstruction, etc.). This makes the method compatible with multiple existing systems, as to potentially enhance them for higher-level behaviours.

The detection of handover intention of the two predictors has been tested both on data recorded from the previous experiments and with a real setup, and has shown very promising results.

5 Projects and collaborations

The work presented has been supported by the European Commission under the following projects.

HBP - The Human Brain Project (HBP) is one of the three FET (Future and Emerging Technology) Flagship projects. The HBP provides a framework where teams of researchers and technologists work together to scale up ambitious ideas from the lab, explore the different aspects of brain organization, and understand the mechanisms behind cognition, learning, or plasticity. (Specific Grant Agreement No. 945539)

APRIL - The APRIL project (multipurpose robotics for mAniPulation of defoRmable materIaLs in manufacturing processes) is developing autonomous, dexterous, and market-oriented robot prototypes to innovate the manufacturing of flexible and deformable materials in European enterprises. (Grant Agreement No. 870142)

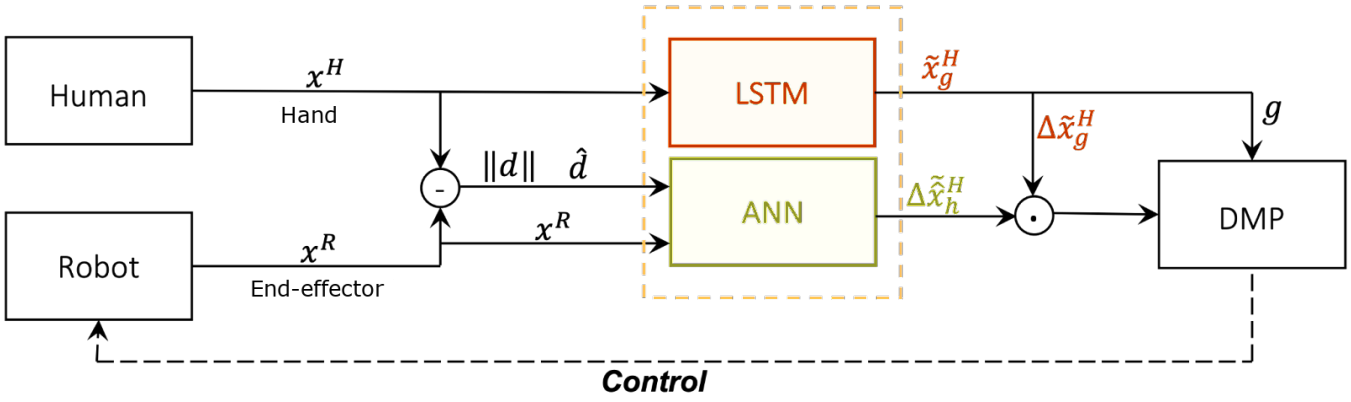
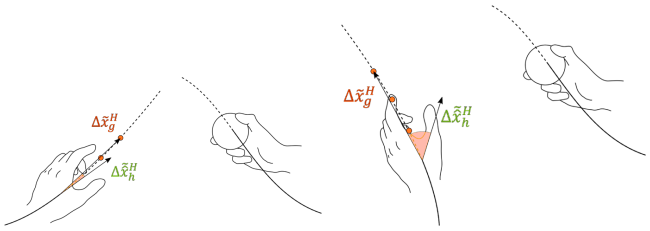


Figure 3: Block diagram of the handover controller based on short-term predictions.



(a) Handover (b) Not handover

Figure 4: Exploiting short-term predictions to detect handover intention.

6 Challenges and perspectives

While the combination of dynamical systems theory and ML is a promising approach for online trajectory generation, there are still some inherent drawbacks to overcome. Mainly, as the trajectory is not planned before the execution, the problem of avoiding "bad" configurations (e.g. reaching joint limits, self-collision, etc.) becomes more difficult. In the same way, it can be more difficult to generate trajectories in cluttered environments. However, these problems are not new to the field, and many proposed solutions already exist. Furthermore, given the relative simplicity of collecting data for such geometrical problems even inside a simulation, the application of ML could improve this kind of online algorithms significantly, while allowing to keep their typical responsiveness and adaptability. Addressing these hindrances will be the focus of our future work on this topic.

Finally, by including preferential feedback, an intuitive way to tune non-linearly correlated parameters is presented. Approaches like this would allow users to tune robot behavior without any expert knowledge. However, BO remains rather greedy, and while it allows for fast learning, it tends to overfit in the short term. Thus, investigation towards optimizing the algorithm, and possibly increasing the number of parameters, is warranted.

References

[Brochu *et al.*, 2010] Eric Brochu, Vlad M. Cora, e Nando de Freitas. A Tutorial on Bayesian Optimization of Expensive Cost Functions, with Application to Active User Modeling and Hierarchical Reinforcement Learning. *arXiv:1012.2599 [cs]*, December 2010. arXiv: 1012.2599.

[Chu e Ghahramani, 2005] Wei Chu e Zoubin Ghahramani. Preference learning with Gaussian processes. In *Proceedings of the 22nd international conference on Machine learning - ICML '05*, pages 137–144, Bonn, Germany, 2005. ACM Press.

[Ijspeert *et al.*, 2002] A. J. Ijspeert, J. Nakanishi, e S. Schaal. Movement imitation with nonlinear dynamical systems in humanoid robots. In *Proceedings 2002 IEEE International Conference on Robotics and Automation (Cat. No.02CH37292)*, volume 2, pages 1398–1403 vol.2, May 2002.

[Ortenzi *et al.*, 2021] Valerio Ortenzi, Akansel Cosgun, Tommaso Pardi, Wesley P. Chan, Elizabeth Croft, e Dana Kuli. Object Handovers: A Review for Robotics. *IEEE Transactions on Robotics*, 37(6):1855–1873, December 2021. Conference Name: IEEE Transactions on Robotics.

[Prada *et al.*, 2014] Miguel Prada, Anthony Remazeilles, Ansgar Koene, e Satoshi Endo. Implementation and experimental validation of Dynamic Movement Primitives for object handover. In *2014 IEEE/RSJ International Conference on Intelligent Robots and Systems*, pages 2146–2153, September 2014. ISSN: 2153-0866.