Probabilistic Online Robot Learning via Teleoperated Demonstrations for Remote Elderly Care

Meccanici, Floris; Karageorgos, Dimitrios; Heemskerk, Cock J.M.; Abbink, David A.; Peternel, Luka

DOI
10.1007/978-3-031-32606-6_2

Publication date
2023

Document Version
Final published version

Published in
Advances in Service and Industrial Robotics - RAAD 2023

Probabilistic Online Robot Learning via Teleoperated Demonstrations for Remote Elderly Care

Floris Meccanici[1,2], Dimitrios Karageorgos[2], Cock J. M. Heemskerk[2], David A. Abbink[1], and Luka Peternel[1(✉)]

[1] Cognitive Robotics, Delft University of Technology, Delft, The Netherlands
l.peternel@tudelft.nl
[2] Heemskerk Innovative Technology B.V., Delft, The Netherlands

Abstract. Daily household tasks involve manipulation in cluttered and unpredictable environments and service robots require complex skills and adaptability to perform such tasks. To this end, we developed a teleoperated online learning approach with a novel skill refinement method, where the operator can make refinements to the initially trained skill by a haptic device. After a refined trajectory is formed, it is used to update a probabilistic trajectory model conditioned to the environment state. Therefore, the initial model can be adapted when unknown variations occur and the method is able to deal with different object positions and initial robot poses. This enables human operators to remotely correct or teach complex robotic manipulation skills. Such an approach can help to alleviate shortages of caretakers in elderly care and reduce travel time between homes of different elderly to reprogram the service robots whenever they get stuck. We performed a human factors experiment on 18 participants teaching a service robot how to empty a dishwasher, which is a common daily household task performed by caregivers. We compared the developed method against three other methods. The results show that the proposed method performs better in terms of how much time it takes to successfully adapt a model and in terms of the perceived workload.

Keywords: Learning from Demonstration · Teleoperation · Online Learning

1 Introduction

Due to an ageing society and shortages of the workforce, elderly care at homes and in nursing homes is becoming one of the key societal challenges [7]. To mitigate the lack of caregivers with respect to the increasing numbers of elderly, we see service robots and teleoperation as one of the most promising solutions. We envision deploying multiple robots to various locations, while caregivers can teach and correct them remotely when they get stuck, thus eliminating travel time.

Most of the existing Learning from Demonstration (LfD) is done by humans kinaesthetically guiding the robot on how to perform manipulation tasks [2,8,14], which requires physical presence of the teacher and thus does not fit with our vision. On the other hand, robot teaching can also be done through teleoperation, where the teacher performs demonstration remotely through various interfaces [12,15]. Skills for complex

T. Petrič et al. (Eds.): RAAD 2023, MMS 135, pp. 12–19, 2023.
https://doi.org/10.1007/978-3-031-32606-6_2

manipulation tasks are typically encoded by trajectories such as Dynamic Movement Primitives (DMPs) [13]. However, the deterministic nature of DMPs offers limited generalisation with respect to unpredictable situations. On the other hand, probabilistic trajectories such as Probabilistic Movement Primitives (ProMPs) [9] use a probability distribution inferred from multiple demonstrations for improved generalisation.

A general LfD approach is to train a skill model offline and expect it to perform well afterwards. However, problems arise when the trained model is incorrect, either due to insufficient training or unknown variations that have occurred in the environment. This demands methods that are able to refine the initially trained model, of which the state-of-the-art can be divided into online [1,4,11,14] and active learning [3,8]. In online learning, the operator is able to refine the motion during the execution time, while active learning uses an uncertainty measurement to query the operator for a new demonstration. In care scenarios, we want to adapt the model towards unknown changes in the environment, and since active learning needs to encode the environment to determine the uncertainty in the prediction, these methods are not well suited. On the other hand, in online learning, the operator has the role to intervene when the model needs to be adapted, and is, therefore, better suited in an unknown environment. State-of-the-art online learning methods mainly use kinaesthetic teaching to refine the motion [4,14], which are not suitable for the remote elderly care service.

Teleoperated online teaching is preferred in the home-care scenario, because it does not require an operator to be constantly physically present at the robot and thus enables easy switching between teaching multiple robots at different locations. The methods in [10,15] enabled teleoperated teaching, however, the learning process was done offline. The method in [12] enabled online teaching by encoding the variance indirectly through a separate deterministic stiffness trajectory. However, encoding variance indirectly through deterministic methods makes the skill model less rich and less generalisable. The method examined in [6] used probabilistic encoding in combination with teleoperated teaching, however, they offered no online refinement mode for specific corrections/updates of the trained trajectory. Some online teleoperated teaching methods used a shared control approach to arbitrate the level of autonomy between the robot skill and the human teacher [1,11]. Since we want humans to be readily in control when needed and to quickly adapt the motion in order to deal with unknown/changing parts of the environment, these types of methods are also not well suited for our problem. Therefore, the existing literature is missing a teleoperated online learning method that allows for quick refinements of the executed predicted motion of a probabilistic skill model to deal with changing environment.

To close this gap we develop a teleoperated online learning approach with a novel refinement method that can quickly adapt the executed predicted motion using a haptic device. The method is based on ProMP, where the detected object is an input and the probabilistic trajectory is an output. To analyse the usability of the proposed method, we conduct an extensive human factors experiment.

2 Method Design

To achieve the desired teleoperated online learning, we built an LfD framework by using ProMPs (see Fig. 1). Inspired by work in kinaesthetic teaching [4], we conditioned

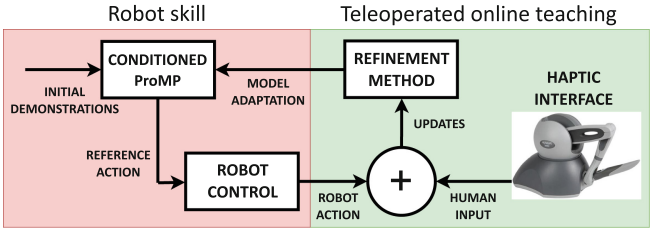

Fig. 1. Overview of the proposed framework. Initial demonstrations are used to train a skill model offline using conditioned-ProMPs. This model generates an action, which is adapted online using the proposed novel teleoperated online refinement method, and is used to updated the model.

ProMPs on an external state variable $\boldsymbol{s}$ (in our case different object positions and initial robot poses). After using the conditioned-ProMPs to train an initial skill model offline, the predictions were refined in an online manner by the proposed novel teleoperated online refinement method. The refinements were then used to update the skill model.

To enable the operator to effectively adapt the executed trajectory online via teleoperation, we developed a novel online refinement method, where both visual and haptic feedback was provided to the human operator. The online learning started with an initial skill model from ProMPs that produces a desired reference trajectory $\boldsymbol{\tau}_d$, which was used by the robot controller to produce the actually executed trajectory $\boldsymbol{\tau}_a$ (in the case of ideal controller $\boldsymbol{\tau}_a = \boldsymbol{\tau}_d$). Then the operator was able to adapt this trajectory in a shared control manner (Fig. 1), resulting in human refinement trajectory $\boldsymbol{\tau}_{hr}$. To do this, the operator moved the master (haptic) device in the desired adaptation direction and, while doing so, he/she felt a force proportional to the magnitude of the refinement.

Rather than adapting the position at the current time step, we adapted the position of the executed trajectory at the next time step by using the current master position (Fig. 2). To achieve this, we calculated the position vector in the next step $i+1$ relative to the current step i (called the $i+1$ and i frame, respectively) as ${}^i\boldsymbol{p}_{i+1} = {}^i R_b({}^b\boldsymbol{p}_{i+1} - {}^b\boldsymbol{p}_i)$, where position vector is $\boldsymbol{p} = [x,\ y,\ z]^T$ and ${}^iR_b$ is a rotation matrix between the base frame and the current frame. The master position was normalised so that the zero position was approximately in the middle of the workspace, which can be seen in Fig. 2. To get the position of the human intervention (green dot in Fig. 2) combined with the reference position, we added this normalised master position ${}^i\boldsymbol{p}_m$ to ${}^i\boldsymbol{p}_{i+1}$ as ${}^i\boldsymbol{p}_{i+1,new} = {}^i\boldsymbol{p}_{i+1} + {}^i\boldsymbol{p}_m$. The end-effector position expressed in the base frame, which was used for the robot control, was obtained by ${}^b\boldsymbol{p}_{i+1,new} = {}^b\boldsymbol{p}_i + {}^b R_i({}^i\boldsymbol{p}_{i+1,new})$.

After the operator created $\boldsymbol{\tau}_{hr}$, which is ${}^b\boldsymbol{p}_{i+1,new}$ at every time step, the reference $\boldsymbol{\tau}_d$ was updated using $\boldsymbol{\tau}_d^{\text{new}} = \boldsymbol{\tau}_d^{\text{old}} + \alpha(\boldsymbol{\tau}_{hr} - \boldsymbol{\tau}_a)$, where $\alpha \in [0, 1]$ defined how much the difference between $\boldsymbol{\tau}_{hr}$ and $\boldsymbol{\tau}_a$ changes $\boldsymbol{\tau}_d^{\text{old}}$. This value can be adapted based on the confidence of the operator in its refinement, but in this application, α is set to 1, as in [4]. This loop continued until the prediction was successful. Visual feedback to the operator was provided through the wrist and head camera views, and additionally, the point cloud of the environment was shown, enabling visualisation of the execution ($\boldsymbol{\tau}_a$) and refined trajectory ($\boldsymbol{\tau}_d^{\text{new}}$) with respect to the environment.

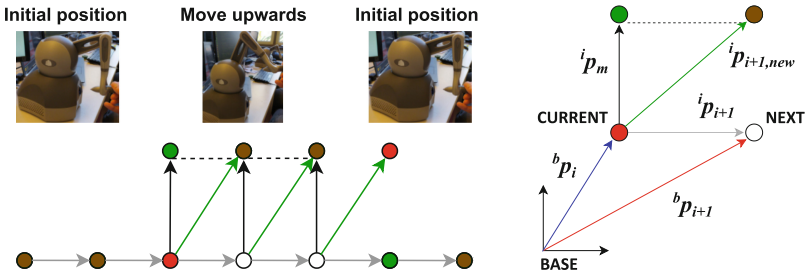

Fig. 2. Illustration of corrections to the executed trajectory. The green and red dots represent the master and slave (remote) robot positions, respectively. When the master and slave are in the same position, the dot has a brown colour. The white dots represent the reference trajectory, where the executed trajectory converges if the operator does not apply any input. (Color figure online)

After refining the trajectory, $\boldsymbol{\tau}_d^{\text{new}}$ was used to update the conditioned-ProMPs. First, new ProMP weights $\boldsymbol{w}$ were calculated from $\boldsymbol{\tau}_d^{\text{new}}$, after which a vector $\boldsymbol{x}$ was created by appending condition $\boldsymbol{s}$ (i.e., different object positions and initial robot poses) to ProMP weights $\boldsymbol{w}$ as $\boldsymbol{x} = [\boldsymbol{w}^T, \boldsymbol{s}^T]$. $\boldsymbol{x}$ was then used with Welford's method for updating the mean and covariance incrementally, which means that one data sample was used [4]. This method was able to quickly update the mean and covariance, without having to store all previous data. To align demonstrated trajectories with different time durations, we employed Dynamic Time Warping.

3 Human Factors Experiment

Eighteen participants volunteered for this experiment and signed an informed consent form prior to participation. The participants were asked about their gaming and teleoperation experience since this information was relevant to the analysis. The research was approved by the Human Research Ethics Committee of Delft University of Technology.

3.1 Experiment Setup

The experiment setup is shown in Fig. 3. The goal of the human factors experiment was to compare the proposed method against three other methods for adapting a trained skill model. These methods were obtained by a combination of two learning mechanisms (online or offline learning) and two teaching devices (haptic interface or teach pendant), leading to four experimental conditions. In online learning, the refinement was done during the trajectory execution, while in offline learning the trajectory had to be recreated as a whole prior to execution. The developed online learning combined with a haptic stylus interface (Geomagic Touch) is the proposed method (*OnStyl*). The teach pendant is still the most common device used in the industry and was emulated with a generic PC keyboard and employed either in offline *OffKey*) or online (*OnKey*) learning. The teleoperated offline LfD was performed using a haptic stylus interface (*OffStyl*), where demonstrated trajectories were recorded and then ProMP weights were recalculated offline. Before starting each experiment condition (method), the participants went

through familiarisation trials to minimise the human learning effect on the results. We counterbalanced the effects of the order in which the participants are presented with different experiment conditions by using a balanced Latin Square experimental design.

Fig. 3. Experiment setup with the teaching devices (purple), as well as graphical user interface (GUI) for visual feedback and a monitor to fill in the subjective questionnaire (NASA TLX).

We hypothesised that *the proposed method, which is a combination of online learning and haptic stylus, has the lowest refinement time and workload*. To test the hypothesis, we used an experiment task of moving dishes out of a dishwasher, which is one of the common daily tasks performed by caregivers. The complexity of this task comes from a cluttered and unpredictable environment with many environmental constraints. The participants had to operate a service humanoid robot to teach and refine the skill model for the given task in a simulated environment. Each participant had to adapt the model for one object, where its positions differed among the participants. Three initial models (i.e., reaching trajectories) had to be adapted with each of the four methods per participant. For each model, the participant had eight attempts to update the model, and per update ten refinement attempts. A refinement was defined as the creation of an adaptation of the trajectory (or previous refinement). If the participant failed to adapt the model within these attempts, we classified it as a failure and excluded this specific participant data from the refinement time hypotheses testing.

We analysed the performance in terms of how much time it takes the human users to successfully adapt a model (refinement time). Additionally, we analysed the perceived workload by the participants, which was evaluated using NASA TLX [5]. Finally, after the experiment, each participant was asked which method they liked the most and why. We analysed the significance of the results with a statistical test. Because both the refinement time and the workload data were not normally distributed, we used Wilcoxon signed rank test instead of a one-sided t-test. The level of significance was set to 0.05.

3.2 Results

An example of reaching for four different object positions is shown in Fig. 4. Refinements were only necessary for the first two object positions, after which the model was able to generalise for the other two object positions without further refinements.

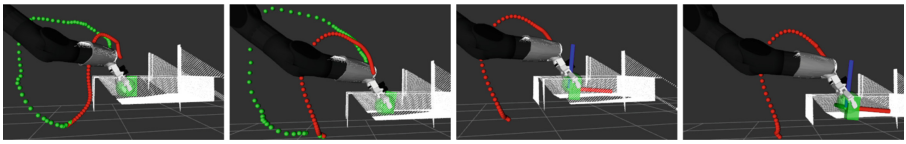

Fig. 4. Reaching for four different object positions inside the dishwasher. The executed (red) and refined (green) trajectories are indicated with point clouds. (Color figure online)

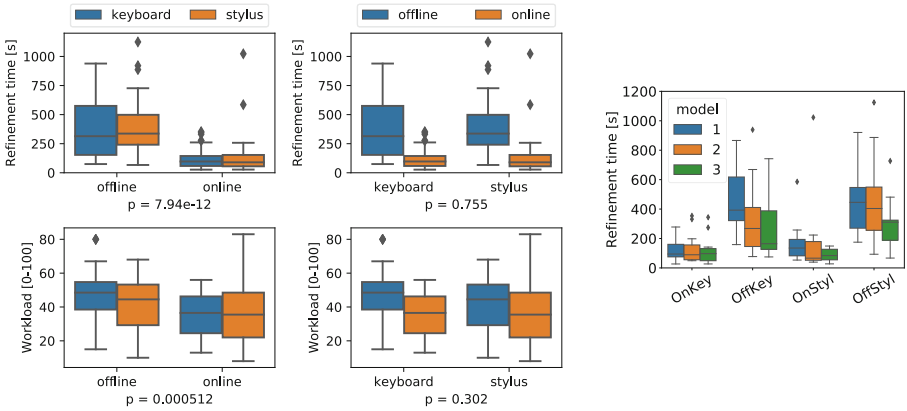

Fig. 5. Left four graph: main results. Right graph: refinement time per model. Blue, orange and green colours represent models 1, 2 and 3, respectively. The boxplots report the median (M), first quartile (25) and third quartile (75) of the refinement time and workload. The statistical significance of the difference between conditions is indicated by p-values below the graphs. (Color figure online)

Out of 18 participants, six failed either in the training or to adapt a model in one of the methods. This means that in total 12 participants were used for the data analysis of the refinement time hypothesis, thus 36 models were evaluated per each method (experiment condition). Figure 5 shows the experiment results, where the distribution of this data and the refinement time per model and per method are depicted. The online methods performed much better compared to the offline methods in both refinement time and workload (i.e., lower values indicate better in both cases). The difference in refinement time was statistically significant both online ($p < 0.001$) and offline ($p < 0.001$). However, there was no statistically significant difference in refinement time ($p = 0.755$) and workload ($p = 0.302$) between the keyboard and stylus.

No Wilcoxon signed rank test can be performed on any background parameter except the gaming experience since the sample sizes are not equal. The participants with high gaming experience perceived a lower workload for all methods compared to having low gaming experience. The difference was statistically significant ($p = 0.048$). On the other hand, there was no statistically significant difference between having high or low gaming experience in terms of the refinement time ($p = 0.189$).

The results of the questionnaire regarding the method preference showed a strong preference for the proposed method (*OnStyl*), where 8 out of 18 participants voted for it.

The second most popular was the online learning method using teach pendant (*OnKey*) with 4 votes. The least popular were offline methods (*OffKey* and *OffStyl*), both receiving 3 votes each.

4 Discussion

Results show that, by using the developed method, human operators are able to adapt an initially trained model to account for unknown variations: goal deviation and an unforeseen obstacle. The method was able to update the model for different object positions. It should be noted that the amount of conditions to refine is dependent on the number of initial trajectories used to train the initial model, and how complex the motion is.

The significant difference in refinement time and workload in favour of online methods can most likely be attributed to the operator only performing small adjustments to the executed trajectory in online learning, instead of completely teaching a new one as in offline learning. This means that online methods are inherently faster in creating a single refinement and a lower amount of input from the operator is needed, possibly resulting in a lower workload as well. Interestingly, there is also a lower variability in refinement time in the online compared to offline methods (Fig. 5).

Another influence on the refinement time is the task execution strategy, which tended to change within and between the methods. The within change can be explained by the right graph of Fig. 5, where the refinement time decreases as a function of the model number. This gives an indication that either the operator has a constant strategy in its mind and is figuring out how to translate this strategy to a demonstration using the specific method, or the strategy changes and the operator is figuring out what strategy works best. Since the offline methods have a higher slope in the medians of the refinement time compared to online methods and the method exposure was counterbalanced, it seems more reasonable that the participants had a more or less constant strategy, but had more trouble using the offline methods to convey this strategy. This suggests that more operator training is needed to use offline methods effectively.

No significant difference in the type of teaching device (stylus or keyboard) in both refinement time and workload could be because the intuitiveness of the interface is person dependent. Some participants reported *OnKey* to be easy and less sensitive than *OnStyl* and preferred the limited degrees of freedom (DoF). On the other hand, others felt that they could easily press the wrong buttons on the teach pendant and found it hard to figure out the axes. Another explanation could lie in the complexity of the corrections, wherein in the examined cases the corrections were mostly in one DoF, i.e., movement along the vertical axis to avoid the dishwasher basket. Increasing the complexity of correction might favour *OnStyl* but additional study is required to provide that insight.

Another explanation for finding no significant difference between using the stylus and keyboard is that there is a difference in the implementation of the offline methods. In *OffKey* the participants could take their time, as it did not matter what was done between the waypoints as long as the waypoints are correctly specified. *OffStyl*, on the other hand, continuously tracked the demonstrated motion, and when the operator made a small mistake this could translate into an unsuccessful demonstration more easily. Therefore, *OffStyl* could be implemented similarly as the teach pendant, where the operator can specify waypoints and interpolate between them. This is expected to

show statistical results in favour of the stylus since the definition of offline will be more similar between the stylus and the keyboard implementation.

The proposed method was tested on one representative task in elderly care, i.e., emptying a dishwasher. In future, other common daily tasks found in elderly care should be examined, such as setting the table for breakfast and cleaning it afterwards. The method itself could be improved by enabling robots to also learn cognitive skills.

References

1. Abi-Farraj, F., Osa, T., Peters, N.P.J., Neumann, G., Giordano, P.R.: A learning-based shared control architecture for interactive task execution. In: 2017 IEEE International Conference on Robotics and Automation (ICRA), pp. 329–335. IEEE (2017)
2. Calinon, S., Sardellitti, I., Caldwell, D.G.: Learning-based control strategy for safe human-robot interaction exploiting task and robot redundancies. In: 2010 IEEE/RSJ International Conference on Intelligent Robots and Systems, pp. 249–254. Citeseer (2010)
3. Conkey, A., Hermans, T.: Active learning of probabilistic movement primitives. In: 2019 IEEE-RAS 19th International Conference on Humanoid Robots (Humanoids), pp. 1–8. IEEE (2019)
4. Ewerton, M., Maeda, G., Kollegger, G., Wiemeyer, J., Peters, J.: Incremental imitation learning of context-dependent motor skills. In: 2016 IEEE-RAS 16th International Conference on Humanoid Robots (Humanoids), pp. 351–358. IEEE (2016)
5. Hart, S.G., Staveland, L.E.: Development of NASA-TLX (task load index): results of empirical and theoretical research. In: Advances in Psychology, vol. 52, pp. 139–183. Elsevier (1988)
6. Havoutis, I., Calinon, S.: Learning from demonstration for semi-autonomous teleoperation. Auton. Robot. **43**(3), 713–726 (2019)
7. Kovner, C.T., Mezey, M., Harrington, C.: Who cares for older adults? Workforce implications of an aging society. Health Aff. **21**(5), 78–89 (2002)
8. Maeda, G., Ewerton, M., Osa, T., Busch, B., Peters, J.: Active incremental learning of robot movement primitives. In: Conference on Robot Learning, pp. 37–46. PMLR (2017)
9. Paraschos, A., Daniel, C., Peters, J.R., Neumann, G.: Probabilistic movement primitives. In: Advances in Neural Information Processing Systems, pp. 2616–2624 (2013)
10. Pervez, A., Latifee, H., Ryu, J.H., Lee, D.: Motion encoding with asynchronous trajectories of repetitive teleoperation tasks and its extension to human-agent shared teleoperation. Auton. Robot. **43**(8), 2055–2069 (2019)
11. Peternel, L., Oztop, E., Babič, J.: A shared control method for online human-in-the-loop robot learning based on locally weighted regression. In: 2016 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), pp. 3900–3906. IEEE (2016)
12. Peternel, L., Petrič, T., Oztop, E., Babič, J.: Teaching robots to cooperate with humans in dynamic manipulation tasks based on multi-modal human-in-the-loop approach. Auton. Robot. **36**(1–2), 123–136 (2014)
13. Saveriano, M., Abu-Dakka, F.J., Kramberger, A., Peternel, L.: Dynamic movement primitives in robotics: a tutorial survey (2021). arXiv preprint arXiv:2102.03861
14. Saveriano, M., An, S.i., Lee, D.: Incremental kinesthetic teaching of end-effector and null-space motion primitives. In: 2015 IEEE International Conference on Robotics and Automation (ICRA), pp. 3570–3575. IEEE (2015)
15. Schol, J., Hofland, J., Heemskerk, C.J., Abbink, D.A., Peternel, L.: Design and evaluation of haptic interface wiggling method for remote commanding of variable stiffness profiles. In: 2021 20th International Conference on Advanced Robotics (ICAR), pp. 172–179. IEEE (2021)