Towards Industrial Robot Learning from Demonstration

Wilson K. H. Ko
ETH Zürich
Zürich, Switzerland
wko@student.ethz.ch

Yan Wu
A*STAR Institute for
Infocomm Research
Singapore
wuy@i2r.a-star.edu.sg

Keng Peng Tee
A*STAR Institute for
Infocomm Research
Singapore
kptee@i2r.a-star.edu.sg

Jonas Buchli
ETH Zürich
Zürich, Switzerland
buchlij@ethz.ch

ABSTRACT
Learning from demonstration (LfD) provides an easy and intuitive way to program robot behaviours, potentially reducing development time and costs tremendously. This is especially appealing for manufacturers interested in using industrial manipulators for high-mix production, since this technique enables fast and flexible modifications to the robot behaviours and is thus suitable to teach the robot to perform a wide range of tasks regularly. We define a set of criteria to assess the applicability of state-of-the-art LfD frameworks in the industry. A three-stage LfD method is then proposed, which incorporates human-in-the-loop adaptation to iteratively correct a batch-learned policy to improve accuracy and precision. The system will then transit to open-loop execution of the task to enhance production speed, by removing the human teacher from the feedback loop. The proposed LfD framework addresses all criteria set in this work.

General Terms
Computer Applications; Industrial control

Author Keywords
Robot Learning from Demonstration; Imitation Learning; Industrial Robots; Industrial Robotics; Automation; Human-Robot Interaction; Intuitive Teaching

INTRODUCTION
In manufacturing, *mix* refers to the variety of products while *volume* refers to the quantity [6]. Supply chains can be categorised by a combination of mix and volume. For low-mix low-volume tasks, such as production of luxury items, manual labour is used while for low mix high volume industrial robots are the opted approach.

HAI 2015, October 21 - 24, 2015, Daegu, Kyungpook, Republic of Korea

ACM 978-1-4503-3527-0/15/10 ...$15.00
DOI: http://dx.doi.org/10.1145/2814940.2814984

Low-mix high-volume settings are mostly controlled environments in which manipulators are required to perform predefined, highly repetitive tasks. Highly skilled workers are required to set up the environment and program the exact robot sequence, which take a long lead-time. This becomes problematic for high-mix low-volume manufacturers which are common in small and medium enterprises (SMEs), where each produced item is highly customized to a small number of customers. Furthermore, a general shift from mass production to mass customization is observed in the past few years, which requires the setup and the robot tasks to be reconfigurable [12].

Learning from demonstration (LfD) is an extensively studied topic in human-robot interaction and skills transfer. It is an interesting solution to this new challenge as it offers quick and easy teaching of different tasks [3]. This paradigm offers a unique niche area in high-mix low-volume production where it is more important for robots to be able to cope with a wide variety of settings than to be highly repetitive. Most work on LfD are for learning general skills and do not specifically cater to industrial purposes [1, 3]. However, apart from the limited LfD work targeted at industrial applications, some other work is potentially applicable to them. In the following section, we detail a list of industrial criteria to assess applicability of LfD frameworks to industrial purposes. We then discuss state-of-the-art work according to these criteria. Based on the assessment, an LfD model that addresses the criteria is proposed.

INDUSTRIAL CRITERIA
Since articulated robots are often employed to do fast and repetitive manufacturing tasks, they need to fulfil requirements specified by the International Organization for Standardization (ISO) to ensure safe [8] and qualitative task performance [7]. However, there are scenarios where these standards might be violated. This means that, when developing an LfD algorithm for industrial robot control, these standards still need to be considered. Aside from the requirements instilled by ISO, criteria regarding the performance and ease of use of LfD algorithms need to be considered when comparing the methods. These are often practical requirements. The

remainder of this section will elaborate on the criteria we set forth.

Accuracy and repeatability

Accuracy refers to how close a given value is to the goal or true value. In the context of industrial manipulators, this means how close the manipulator is able to follow a desired kinematic or dynamic trajectory or how close it gets to its goal. *Repeatability* and *precision* are synonyms and refer to how well a manipulator is able to reproduce a certain action.

Adaptability and generality

In this work, the terms adaptability and generality are used interchangeably and refer to the degree to which an algorithm can be applied to learn different tasks, or to perform a learned task in different settings without changing the system. This also includes applicability to different systems without changing much. Having high adaptability removes the need for experts to reprogram the manipulator, but only requires the operator to re-teach the robot.

Learning fatigue

High learning fatigue in this work means that it is hard to teach the robot a new task or that re-teaching needs to be done often for the same task in different settings. Note that this is not the same as *adaptability*. A system with high adaptability but also high learning fatigue is still undesirable for operators. Although no expert knowledge is needed in this case, it is time consuming having to repeat demonstrations.

System complexity

By system complexity, we refer to the size of the system architecture. This means that system complexity is about the number of sub-components the system consists of, i.e. use of additional sensors, reliance on detection and recognition algorithms etc. High system complexity means that the system relies heavily on these additional sensors to be able to extract data for learning or execution.

Production speed

For industrial applications, it is desirable for the robot to have fast execution in order to optimise task efficiency. As mentioned before, industrial robots are required to be fast by ISO, but the LfD algorithm should make full use of this potential speed during reproduction in order to be useful.

LFDS IN INDUSTRIAL APPLICATIONS

LfD methods encapsulate algorithms which learn how to perform a task, given some demonstrations of how to do the task. The operator provides the learner demonstration examples, from which the learner approximates the underlying general function to be used for reproduction of the task. Here we evaluate existing LfD frameworks using the aforementioned criteria and assess their effects on these criteria.

Effects on accuracy and precision

Most LfD works evaluate accuracy and precision as an integrated metric [9, 5, 4, 14], while others do not present information on accuracy at all [13]. In [10], the authors present the tracking errors for position and force profiles under different circumstances, attributing the learning inaccuracies to the robot gravity compensation module.

To ensure high production precision, the type of robot executing the tasks plays a vital role. [13] uses the 6-axis FANUC robots to execute the gesture-based LfD tasks.

Precision can also be addressed by parameter-tuning. In [9], higher precision is achieved by fine-tuning the reward function for PI^2. The robot is able to open doors and grasp pens with a 100% success rate, showing both high precision and accuracy. A similar approach is presented in [5] to adapt arm stiffness when enhanced precision is required.

High precision is achieved by also adjusting the arm stiffness according to the variance in the training data [10]. Precision of the reproduction stage is improved by sampling from the learnt task parameters during the exploration phase [4]. More precision could be achieved by increasing the parameter size, although this is a trade-off with the number of parameters to learn and thus learning fatigue. However, improvement in precision and accuracy is still required for these algorithms [14].

Effects on adaptability and generality

Adaptability can be increased by ensuring that the same algorithm can be used to learn many different tasks [9, 4, 13], or by ensuring that a certain learned task can be executed under different circumstances [5, 14], or both [10].

During the learning stage, high adaptability can be achieved by using a generic interface which can be used to teach multiple types of robots. This is achieved by doing as little programming as possible on the robots, so the gesture recognition system becomes more generic [13]. Moreover, this interface can be used to demonstrate a wide range of movements, enabling the robot to learn a wide variety of tasks if used in combination with a policy derivation algorithm.

Different tasks can also be learned by redesigning the reward function for RL algorithms [9]. The same RL algorithm enabled [5] to apply the learning task to different robots. However, although it is able to learn different tasks, changes in the learning setup are generally needed for RL. On another note, due to the exploring nature of RL, these algorithms can achieve high adaptability in different, unknown circumstances.

Similar to [5], adaptive stiffness is achieved in [10]. Furthermore, evaluations of the force and position controller show that the algorithm is able to cope with different circumstances. Generality is increased by using a task-parameterized model during learning in [4].

Effects on learning fatigue

In order to reduce learning fatigue, demonstrating should occur in an intuitive and easy fashion. *Stipanvcic et al* acknowledge this point and state that their gesture language is rather hard to learn and to apply, making it hard to teach a task [13]. Although different, RL algorithms [5, 14, 9] face a similar problem in choosing a good initial policy and reward function, which may not be straightforward. Without a good initial policy, RL algorithms may not converge well, or end up taking more time. Without a good reward function, the agent might not learn the task well. The latter problem could be solved by using Inverse Reinforcement Learning (IRL) [11], although with this method it still holds that demonstrations should be easy to show.

Learning fatigue can also be reduced by recycling useful knowledge [14]. The authors state however, that satisfying results can only be obtained by providing qualitative demonstration data, meaning that learning fatigue can be high if the demonstrations are qualitatively poor. Thus, being able to cope with bad demonstrations autonomously might reduce learning fatigue even further. In [4], the aim was to actively reduce the number of demonstrations needed by using an autonomous exploration algorithm. By allowing the task parameters to change for new situations, it is not longer needed to re-teach the system for these situations.

Effects on system complexity

Most of the mentioned work employ at least one additional sensor such as IMU's or cameras. We believe that low system complexity is achieved when the bare minimum of equipment and tools is used to learn or perform a task. We kept into consideration that having at least one additional sensor is common for feedback control [9, 13, 14].

An exception is [10], as an haptic device is used which is more uncommon than a standard camera or IMU, since haptic devices are relatively state-of-the-art and expensive. Moreover, the haptic device plays a considerable role in the teaching procedure, hence reliability on this external system is rather high. Works with low system complexity usually rely on proprioceptive sensors only which are already included in the industrial robot [5, 4].

Effects on production speed

We observe that none of the mentioned work address production speed, or provide no information about it. Therefore, we were unable to assess any of the discussed work on production speed. However, as mentioned, this criterion is of importance in industrial applications, so we consider production speed of LfD frameworks an open area in research.

LAP - A THREE-STAGE LFD FRAMEWORK

We discussed several LfD approaches for industrial applications based on the criteria set in this work. All of them focus on at least one of the criteria, but few of them address all the important aspects for LfD to be relevant in the industry. In the following section, we present a novel LfD method which addresses all the mentioned criteria. The open area of production speed in LfD will be the focus of this work.

Overview

In the literature, LfD has been applied to reinforcement learning (LfD-RL) where task demonstrations are used to reduce the RL search space by providing an initial estimate of the state-space [2]. This enables the system to quickly generalise and converge to an accurate and precise solution.

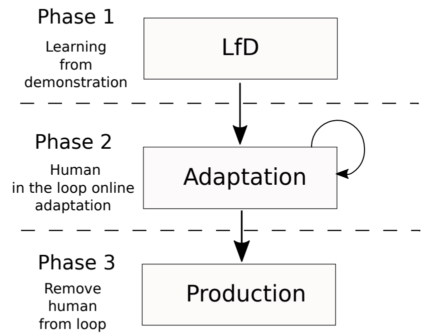

Figure 1. Proposed three-stage LfD method. Adaptation phase is iterative with human corrections until high precision and accuracy is achieved.

In this work, we present a novel LfD concept of Learning-Adaptation-and-Production (LAP), a three-stage framework to address the challenges in industrial LfD as shown in Figure 1. Similar to the LfD-RL approach, batch learning will be carried out in the first stage to generalise the demonstrated tasks. Humans will be used to adapt the robot's performance in the second stage. Once the learned outcome attains accuracy and precision for open-loop execution, the production stage can be executed for mass production.

The Learning Stage

The design of the learning stage takes into consideration the adaptability, learning fatigue and system complexity. The system will learn low-level task profiles, such as task-space force and pose trajectories by using a suitable learning algorithm that can generalise with reasonable accuracy from a limited number of demonstrations. The reason to perform task-space learning is to minimise adaptability issues, as the learned outcome is an encapsulation of the robot-independent task which can be transferred to any other system easily. We use DMP to learn these low-level trajectories to allow for generality since DMPs can be easily adjusted by changing the learned weights. A regression technique will be used to cope with few demonstrations. Kinaesthetic teaching will be used to minimise the effect of the correspondence problem [1].

The Adaptation Stage

Instead of reinforcement learning used in the LfD-RL framework, which uses the robot to actively search through its state-space, we expedite the learning process by using a human-in-the-loop approach. This is because most LfD-RL work which can change parameters autonomously on-the-fly [5, 4, 10] require complex probabilistic algorithms, additional sensors and environmental constraints to evaluate the performance which adds to system complexity and reduces adaptability. With a human operator, the high level evaluation can be carried out efficiently. Thus, the adaptation stage will include the human operator as a feedback and correction mechanism for the robot in an attempt to reduce system complexity and learning fatigue. During this phase, the robot will reproduce the learned task at a low speed with compliance ready for human intervention at any time. The intervened segments of the tasks will be updated online to reduce learning fatigue. This iteration process will be carried out until satisfactory accuracy and precision are achieved and terminated by the human. The task will then be re-planned and optimised for open-loop execution without compromising accuracy and precision.

The Production Stage

The last stage is the open-loop production stage which is robot-dependent on accuracy and precision. The human operator will be removed from the loop to ensure high speed production. The learned policy in this phase is refined to a level that no more corrections are needed, so the task can be performed in a feed-forward fashion to increase execution speed. Since no more corrections are needed, the robot should now be able to perform the learned task autonomously at high speeds without human guidance.

Application scenario

This framework can be applied to a wide range of industrial manipulation tasks especially for high-volume and low-mix production jobs. We will use the KUKA LBR iiwa robot as a testing platform to teach the robot a pick-and-place/assemble task. A suitable benchmarking method will be introduced to measure the performance of the LfD model in terms of the mentioned criteria. Ideally, this benchmarking method will also be used to explore various settings of the LfD method (e.g. using different regression methods) to find an optimal configuration.

CONCLUSION

In this work we discussed LfD methods, in particular those which are developed with industrial applications in mind or which are applicable to industrial tasks. We explained the criteria which are important for comparing LfD methods for industrial robots. Using this knowledge, different work was discussed and evaluated for industrial applicability, according to the criteria set in this work. Also, a three-stage LfD approach is proposed where policy corrections are made by including humans in the feedback loop, after which the system converges to feed-forward execution of the task. The proposed approach is based on the criteria set in this work; the method will be tested on the KUKA LBR iiwa.

ACKNOWLEDGEMENTS

This work is supported by the Science and Engineering Research Council (SERC, A*STAR, Singapore) Grant No. 1225100001.

REFERENCES

1. Brenna D Argall, Sonia Chernova, Manuela Veloso, and Brett Browning. 2009. A survey of robot learning from demonstration. *Robotics and autonomous systems* 57, 5 (2009), 469–483.
2. Christopher G Atkeson and Stefan Schaal. 1997. Robot learning from demonstration. In *ICML*, Vol. 97. 12–20.
3. A Billard, S Calinon, R Dillmann, and S Schaal. 2008. Robot programming by demonstration. *Springer handbook of robotics* (2008), 1371–1394.
4. Danilo Bruno, Sylvain Calinon, and Darwin G Caldwell. 2014. Learning adaptive movements from demonstration and self-guided exploration. In *Proc. IEEE Intl Conf. on Development and Learning and on Epigenetic Robotics (ICDL-EpiRob)*.
5. Jonas Buchli, Evangelos Theodorou, Freek Stulp, and Stefan Schaal. 2010. Variable Impedance Control-A Reinforcement Learning Approach. In *Robotics: Science and Systems*.
6. Jiri Chod, Nils Rudi, and Jan a Van Mieghem. 2006. Research Articles Mix , Time and Volume Flexibility : Valuation and Corporate Diversification. 2004, January (2006), 262–282.
7. ISO. 1998. *Manipulating industrial robots – Performance criteria and related test methods*. ISO ISO 9283. International Organization for Standardization, Geneva, Switzerland.
8. ISO. 2011. *Robots and robotic devices – Safety requirements for industrial robots – Part 1: Robots*. ISO ISO 10218-1. International Organization for Standardization, Geneva, Switzerland.
9. Mrinal Kalakrishnan, Ludovic Righetti, Peter Pastor, and Stefan Schaal. 2011. Learning force control policies for compliant manipulation. In *Intelligent Robots and Systems (IROS), 2011 IEEE/RSJ International Conference on*. IEEE, 4639–4644.
10. Petar Kormushev, Sylvain Calinon, and Darwin G Caldwell. 2011. Imitation learning of positional and force skills demonstrated via kinesthetic teaching and haptic input. *Advanced Robotics* 25, 5 (2011), 581–603.
11. Gergely Neu and Csaba Szepesvári. 2012. Apprenticeship learning using inverse reinforcement learning and gradient methods. *arXiv preprint arXiv:1206.5264* (2012).
12. Mikkel Rath Pedersen, Lazaros Nalpantidis, Rasmus Skovgaard Andersen, Casper Schou, Simon Bøgh, Volker Krüger, and Ole Madsen. 2015. Robot skills for manufacturing: From concept to industrial deployment. *Robotics and Computer-Integrated Manufacturing* (2015).
13. Tomislav Stipančić, Bojan Jerbić, Ante Bučević, and Petar Ćurković. 2012. Programming an Industrial Robot by Demonstration. In *23rd DAAAM International Symposium on Intelligent Manufacturing and Automation:: Focus on Sustainability*.
14. M Stoica, F Sisak, and AD Morosan. 2012. Reinforcement learning algorithm for industrial robot programming by demonstration. In *Optimization of Electrical and Electronic Equipment (OPTIM), 2012 13th International Conference on*. IEEE, 1517–1524.