# Review: Learning to Scaffold the Development of Robotic Manipulation Skills

Deepak Raina: 2019MEZ8497

COL-864: Homework-3

## Paper Summary

The paper proposes an approach to automatic learning of contact-rich robotic manipulation skills like peg insertion, wrench manipulation and shallow-depth insertion. These skills will make the robot more versatile and robust, while current robots require to adapt to the environment to match their limited abilities. The idea is inspired by human learning principle known as *scaffolding*. Humans leverage the environment to ease various tasks like parents use a training wheel to make their child learn riding a bicycle, industry workers use fixtures to guide their manufacturing procedures. In this paper, physical scaffolding has been used for robotic manipulation skill learning. The two major challenges to this learning approach are:

- High dimensionality of the state and action space: This will drastically increase the learning time as it will allow lot of wrong or undesirable motions. Thus, it will take lot of trials for robot to figure out how to do these tasks correctly.
- Uncertainty from perception and motor control: The inherent noise in the sensors and motors will also cause inaccurate perception and motor control, thereby increasing training time.

To overcome these challenges and achieve robust manipulation, humans exploit contact constraints in the environment. This paper adopted a similar strategy by enabling the robot to modify its environment. They enabled robot to use fixtures which it can place freely in the environment for manipulation skill learning. These fixtures provide hard constraints that limit the outcome of robot actions. Thus they funnel uncertainty from perception and motor control. To learn manipulations skills with this physical *scaffolding* strategy, they proposed a learning system with two loops. In outer loop, robot learns to place fixture in the workspace. In the inner loop, robot learns a manipulation skill and returns a reward to the outer loop after a fixed number of episodes. Thus, outer loop is encouraged to place the fixture in such a way that inner loop reward is maximized. They demonstrated that their framework has sped up the manipulation skill learning for peg insertion, wrench manipulation and shallow-depth insertion tasks. The significant contributions of their work can be listed as follows:

(a) They proposed a learning framework to enable robots to use fixtures to quickly learn complex manipulation tasks

(b) An algorithm has been introduced that improves sample efficiency in bandit problems with continuous action spaces and discontinuous reward functions.

(c) The exciting outcome of this work is that the robot gradually learns to sustain task performance without fixtures, as in the case of human educational learning by *scaffolding*.

## Strengths

They have proposed a learning system consisting of two loops. Inner loop learns fixture pose selection and outer loop learns robot manipulation skill. Briefly, their approach consists of following three modules:

(i) **Fixture pose selection:** The optimal placement of fixture pose has been formulated as *contextual bandit problem*, where *state* is depth image of the scene (i.e. context), *action* is fixture poses (continuous) and *reward* is inner loop accumulated reward (discontinuous). One of the novel contributions of this work is *Smoothed zooming algorithm* for learning discontinuous reward function in continuous action spaces. In this algorithm, the continuous action space is covered with a prefixed number of covering balls with some radius. Then *UCB1* rule is used to select the one covering ball and fixture pose is uniformly sampled from this ball. In this way, they are able to smooth the discontinuous reward function.

(ii) **Robotic skill learning with Fixtures:** This has been formulated as episodic *Reinforcement Learning (RL)* problem. The robot receives an RGB image, takes a action, receives a reward and moves to next state. The reward is sent to the outer loop to guide the fixture pose selection. The objective is to learn a policy and maximize the cumulative reward. They have used a variant of actor-critic algorithm *A3C* to learn the optimal policy.

(iii) **Robotic skill learning without fixtures:** This approach is convincing and shows how robots can benefit from *scaffolding* principle. In this, the hard constraints induced by the fixture has been replaced with a virtual potential field (soft constraints) during training. Potential field has been maintained according to the geometry of fixture sensed by depth camera. The training starts from optimal policy obtained with fixtures. Then the fixture is gradually removed to transfer the skills with fixtures to skills with virtual potential field.

**Results:** To test the efficacy of their framework, experiments have been conducted in both real and simulation. The real-world experiments have been conducted with 7-DoF Franka Panda robot arm while PyBullet has been used for simulation. The details regarding the three manipulation task experiments has been described below in the table.

| Task | Role of fixture | Success |
|---|---|---|
| Peg insertion | Prevents the peg from slipping past the hole | If the peg is completely inserted into the hole |
| Wrench manipulation | Provides support for the wrench to stay horizontal | If bolt is rotated by 30 degrees |
| Shallow-depth insertion | Restricts translation and rotation motion of the cuboid into the slot | If the cuboid is horizontal and inside the shallow slot |

The authors have presented convincing results to show the efficacy of their approach:

- **Performance evaluation with fixture:** It has been reported that proper fixture improves both task performance and learning speed for all tasks, whereas a drop in performance has been observed with suboptimal fixtures.
- **Performance evaluation with fixture removal:** The reasonable performance of the robot has been noted down even when the fixture is completely removed and replaced with artificial potential field.
- **Fixture pose verification:** The learned Q-map has been visualized and it has been noticed that largest Q-values correspond to optimal position of fixtures.
- **Generalization:** One of the most convincing results is the generalization ability of framework across varying peg dimensions. This has been achieved using domain randomization, randomize object positions, textures and lightening conditions during training.

# Limitations & Improvements

The following are some of the limitation of this approach:

(a) The model assumes that the shape of the fixture for the task is prefixed. This assumption has worked well for the mentioned tasks, but it might happen that policy learned for the task is not the optimal policy because there might exist a better shape of fixture that will do the task more efficiently.

(b) The outer loop for fixture pose selection is only exploring the pose in two-dimensional space i.e. $(x, y, \theta)$. This will limit the fixture's ability to one specific task only and user will have to first decide the base of fixture for placement. If another robot could have moved the fixture in three-dimensional space, then one fixture can be used for many tasks, as is done by human workers in industrial manufacturing procedures.

(c) While training the robot without fixtures, they replaced the physical fixtures with artificial potential field. This approach does not look very convincing as learning algorithm is still constraining robot motions virtually. This will limit the framework to simple fixtures as generating a repulsive potential field for complex shape fixtures will be very challenging.

The following are some of the suggestions for improvement of this work:

(i) One of the exciting extension to this approach could be to learn the fixture (scaffold) itself for the given task. This can be achieved using the artificial potential field (soft constraint), which can be leveraged during training to facilitate the learning of optimal shape and position of repulsive potential field as fixture for the manipulation task.

(ii) This learning system can potentially be built into a system in which robot is given arbitrary shape fixture and it will automatically explore and learn how to use it for easing the given task. This can be done by using the second robot in outer loop to move the fixture in three-dimensional space. This capability will make the system modular across various manipulation tasks.

(iii) Since the objective of this process is to speed up the learning process, it would be an interesting prospect to include human awareness of of the task during learning. These approaches have been used in literature for speeding up RL methods. In this paper's approach, this can be done by incorporating human guidance along with RL in the inner loop of manipulation skill learning.