Force-based End-to-end Training of a Mobile Manipulator
—Learning Human Following from Applied Force—

Joshua SUPRATMAN, Student Member, Chiba Institute of Technology, s1426071bl@s.chibakoudai.jp
Yasuo HAYASHIBARA, Member, Chiba Institute of Technology
Kiyoshi IRIE, Member, Chiba Institute of Technology

In this paper, we present a novel method for inducing a mobile robotic manipulator through human force using an end-to-end approach. We applied deep reinforcement learning to train our robot to estimate human intention using the error and the error rate of change for each joint in the manipulator as input state. We used reward method based on stimulus using the displacement between the manipulator’s joint positions when no load is received and the current joint positions. The larger the displacement is the larger the stimulus which gives a negative reward. We successfully trained the robot to respond appropriately to an external force.

Key Words: Human Robot Coordination, Intention Estimation, Reinforcement Learning, Mobile Manipulator

1. Introduction

1.1 Background
Robots and human working in the same environment are increasing such as KAWADA Robotics NEX TGAGE [1] and Universal Robots UR Robot [2]. When designing robots that can perform a task near human, one must consider, in addition to safety, how to convey proper information from human to robot. In general, this could be done using an external controller; however, this method is not always necessarily reliable as communicating to the robot is limited to how much information the controller holds. For example, an external controller or a sensor can stop a working manipulator, but if a human subject collided with the manipulator without the controller at hand or triggering the sensor, the manipulator would keep on moving without stopping. In a situation such as these, the controller may not properly convey human intention to the robot.

When humans interact naturally with each other, they have several means to communicate such as visual, auditory, and physical communication. A father holding hands with his child and an owner pulling the leash to guide their dog are exemplary physical communication. In both cases, one side conveys information by applying force and the other receives information through stimulus. Physical communication is thought to be a simpler method of communication comparing to visual and auditory communication as studied by Geldard [3].

Using the observation as described above, the purpose of this research is attempting to mimic human’s physical communication on a robot by attempting to induce a mobile manipulator, as illustrated in Figure 1, through deep reinforcement learning.

![Figure 1 Illustration of research concept.](image)

1.2 Related Works
There are many studies related to human-robot coordination through physical communication. One of the most commonly known applications is Kosuge et al.’s MS Dance robot [4], which was developed with the goal of human-robot coordination with physical interaction. Predicting human intention through physical communication is generally done where the robot detects and recognizes human intention and plan behavior through human optimization, as studied by Hayashibara et al. [5] and Li et al. [6]. We called this the conventional method.

Many studies use machine learning, to realize human-robot coordination. Modares et al. [7] uses the conventional method but finds optimal parameters using reinforcement learning. Similarly, Ikemoto et al. [8]’s work uses unsupervised learning to obtain optimal parameters by interacting with the human during training. In another related work, Vogt et al. [9] introduce a system for robots to learn to interact with human but through human-human demonstration.

In recent studies, an alternative method to the conventional method was introduced using machine learning as an end-to-end approach. Instead of optimization through human criteria as done in the conventional method, the end-to-end method allows the robot to self-optimize maximum performances. In Bojarski et al.’s work [10], using the end-to-end approach, the car learns to drive through deep learning by directly taking visual observation as input and directly output the car steering. In another related work, Levine et al. [11] work use deep reinforcement learning to achieve end-to-end training.

In contrast to the above research, our work uses deep reinforcement learning to take an end-to-end approach to realize human-robot coordination. Instead of manually defining processing steps, our robot learns to self-optimize all processing steps by interacting kinesthetically with the human. In this way, in addition to robot handling situation outside of planned programming, non-specialist can introduce desire intention without having to explicitly program the process. We believe that this leads to higher acceptance of robots in society.

1.3 Contribution
Our research developed a novel method of human-robot coordination through physical communication by applying an end-to-end approach. We are not aware of a similar system using end-to-end approach for human-robot coordination. We experimentally confirmed that a mobile manipulator can learn to respond appropriately to an external force using only the error and the error rate of change for each joint.

2. Method of Measuring Force
There are two methods of measuring the external force applied to the manipulator:
1) Using force sensor to measure the applied force.
2) Estimating applied force from torque applied to the joints.

Directly measuring applied force using force sensor can provide accurate measurement but cannot measure any force other than where the sensor is mounted. On the other hand, estimating external force from torque applied on the joints is difficult to obtain an estimation must also consider gravity and inertial force of the manipulator.

In order to guide the robot by pulling the manipulator, it is
desirable to induce not only the force applied on the hand alone by on anywhere on the arm. Therefore, in this research, we decided to estimate applied force using torque applied to the joints.

2.1 Estimating torque applied to the joints

There are several methods of measuring the torque applied to the joints such as:
1) Mounting torque sensors on the manipulator
2) Using the current to estimate torque
3) Using the error between the target joint position and the current joint position.

In this research, we use KONDO B3M robot servo motors [12], shown in Figure 2, as manipulator’s joints. The KONDO servo motors are equipped with multiple sensors that can transmit the current state of the servo. As a result of preliminary experiments, we found that the error between the target angle position and the current angle position is closely proportional to the torque. We will use the error from the target position, which is proportional to force, as input to estimate the external force.

(a) B3M-SC-1040-A (b) B3M-SC-1170-A
Figure 2 KONDO B3M robot servo motor used as manipulator’s joints.

3. Induction using human force

To guide the robot using human force, the conventional method is to first measure the force applied to the robot, and then executes a behavior as described in Figure 3.

This method, however, is limited due to several reasons:
1) As mentioned above, estimating external force from torque applied on the joints is difficult to obtain as estimation must also consider gravity and inertial force of the manipulator.
2) Manipulator with multiple degrees of freedom has many states and requires effort to determine the rules for each state to respond appropriately to human intention.

We propose using deep reinforcement learning as an end-to-end approach to guide the robot using human force as described in Figure 4. Comparing to decomposing the problem explicitly and optimizing through human criteria as done in the conventional method, the end-to-end method self-optimizes all processing step simultaneously. This eliminates the need to recognize the external force and plan behavior manually.

External Force
Reinforcement Learning
Action

Figure 4 End-to-end method of inducing a robot.

3.1 Double Deep Q Network

We use Double Deep Q Network (DDQN) reinforcement learning algorithm, proposed by Hasselt et al. [13], which uses the following algorithm

$$Target = r + \gamma Q(s’, argmax(Q(s', a; \theta); \theta^-))$$

where the robot learns a parameterized value function \(Q(s, a; \theta)\) with \(r\) as the reward, \(s\) as the state, \(a\) as the action and \(\gamma\) as the discount ratio, \(\theta\) and \(\theta^-\) is the weight of the current network and the target network respectively.

4. Experiment Condition

We conducted experiments on CIT Brains @Home robot [14]. The robot configuration is shown in Table 1.

We trained our robot with the neural network model shown in Table 2 using Adam optimization algorithm. The input is the error from the target position and the error rate of change for each joint the manipulator. The output gives the Q value for three discrete actions; stop, forward and reverse. The robot executes the action with epsilon-greedy. Table 3 shows the learning parameter.

<table>
<thead>
<tr>
<th>Table 1</th>
<th>Robot configuration</th>
</tr>
</thead>
<tbody>
<tr>
<td>Operating System</td>
<td>Ubuntu 16.04</td>
</tr>
<tr>
<td>Middleware</td>
<td>ROS Kinetic</td>
</tr>
<tr>
<td>Hardware</td>
<td>CIT Brains @home robot</td>
</tr>
<tr>
<td>Actuators for Manipulator</td>
<td>KONDO B3M-SC1170-A, KONDO B3M-SC1040-A</td>
</tr>
<tr>
<td>Actuators for Movement</td>
<td>TF-M30-24-3500-G15R</td>
</tr>
<tr>
<td>Manipulator DOF</td>
<td>5 DOF</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Table 2</th>
<th>Neural network model</th>
</tr>
</thead>
<tbody>
<tr>
<td>Layers</td>
<td>Number of nodes</td>
</tr>
<tr>
<td>Input</td>
<td>10 (error, error change rate)</td>
</tr>
<tr>
<td>Hidden 1</td>
<td>100</td>
</tr>
<tr>
<td>Hidden 2</td>
<td>100</td>
</tr>
<tr>
<td>Output</td>
<td>3 (forward, stop, reverse)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Table 3</th>
<th>Learning parameters</th>
</tr>
</thead>
<tbody>
<tr>
<td>Discount Ratio</td>
<td>0.05</td>
</tr>
<tr>
<td>Learning Rate</td>
<td>0.01</td>
</tr>
<tr>
<td>Batch Size</td>
<td>32</td>
</tr>
<tr>
<td>Memory Size</td>
<td>1000</td>
</tr>
<tr>
<td>Target Network Update</td>
<td>per 5 episodes</td>
</tr>
<tr>
<td>Sampling Size</td>
<td>200 steps per episode</td>
</tr>
</tbody>
</table>

5. Experiment Method

We guide the robot by applying three different forces on the manipulator: push force, pull force and no force. The manipulator is in a fixed posture as shown in Figure 5. The experiments were conducted under two different conditions:
1) Applying force to the end-effector when the robot’s arm is in a fixed position
2) Applying force to anywhere on the robot’s arm when the arm is in a fixed position.
In learning, it is necessary to evaluate whether or not the robot behavior is appropriate to the applied force. We proposed two methods of providing a reward.

5.1 Providing reward explicitly

Judging whether the robot behaves appropriately can be done by explicitly rewarding every action the robot takes. Using an external controller, we provide a reward for every appropriate action the robot takes as illustrated in Figure 6 and described in equation (2). In this paper, we called this the explicit reward method.

\[ \text{Reward} = \begin{cases} 100, & \text{appropriate action} \\ 0, & \text{otherwise} \end{cases} \]  

(2)

5.2 Providing reward through stimulus

When human learn naturally, stimulation impacts learning. For example, when an owner takes a stroll with their dog, the dog receives a small stimulation if they behave appropriately but will receive a strong stimulus, such as a stopping the movement with a strong force if they behave inappropriately.

Using the same observation, we propose a reward method based on stimulus. The robot can judge its own behavior using the displacement between the joint position with no load and the current joint position as correction force as described in equation (3) where \( N \) is the total number of joints.

\[ \text{Reward} = -\sum_{i=1}^{N} (\text{NoLoad} \text{Join}_t_i - \text{Current} \text{Join}_t_i)^2 \]  

(3)

The larger the displacement the stronger the correction force. The robot receives strong correction force for inappropriate behavior and receives minimal correction force for appropriate behavior as illustrated in Figure 7. In this paper, we called this the displacement reward method.

6. Experiment Result

We successfully trained our robot under both explicit and displacement reward method to execute proper behavior when an external force is applied to the manipulator as shown in Figure 8. We first trained our robot under the explicit reward method to confirm that the robot can be guided by human force. After observing that the robot can respond properly to human intention, we trained the robot under the displacement reward method. The training result using the explicit reward method is shown in Figure 9 and 10. The training result using the displacement reward method is shown in Figure 11 and 12.

![Figure 5 Manipulator posture.](image)

![Figure 6 Using an external controller to provide a reward.](image)

![Figure 7 Using displacement between joint position with no load and current joint position to provide a reward.](image)

![Figure 8 Robot responds to human intention.](image)

![Figure 9 Average reward when an external force is applied on the end-effector when trained using the explicit reward method.](image)
incorrectly believe that executing certain action causes no changes in the environment thus leading to poor training as seen in Figure 13. We lowered the control rate from 5Hz to 1Hz which improved the training result as seen in Figure 11 and 12.

Figure 10 Average reward when an external force is applied to anywhere on the arm when trained using the explicit reward method.

Figure 11 Average reward when an external force is applied on the end-effector when trained using the displacement reward method.

Figure 12 Average reward when an external force is applied to anywhere on the arm when trained using the displacement reward method.

We found two crucial factors to successfully train the robot under the displacement reward method,

1) Human subject provide accurate information regarding the current state of the environment

2) Delaying the time between the robot executing an action and observing the next state

When using the displacement reward method, the accuracy of the reward is heavily dependent on the current state of the environment. We, therefore, took extra precaution when applying force on the manipulator. We avoided continuous force (continuing to apply force after the robot behaves appropriately) and interrupting force (applying different force when the robot is executing an action) to provide accurate information of the current environment.

When the robot executes an action based on the observed state, there is a delay between the time the robot interact with the environment and observing the changes in the environment. If the control frequency is too high, the Markov Decision Process assumption can be violated because the robot will observe the next state before the environment actually changes. The robot will incorrectly believe that executing certain action causes no changes in the environment thus leading to poor training as seen in Figure 13. We lowered the control rate from 5Hz to 1Hz which improved the training result as seen in Figure 11 and 12.

Figure 13 Failed training result using the displacement reward method with a control rate of 5Hz.

7. Conclusion

In this research, we aim to realize a robot that can interact with a human through physical communication using an end-to-end approach. We developed a method of inducing a mobile robotic manipulator using human force through deep reinforcement learning: the state observations and reward were given based on the positions of the manipulator's joints. By using this method, we successfully trained our robot to respond appropriately to external forces. Providing accurate information regarding the current state of the environment and avoiding high control frequency were found to be important factors for implementing the method.

REFERENCES