Robot Learning from Human Teachers

Jane Li

Assistant Professor
Mechanical Engineering Department, Robotic Engineering Program
Worcester Polytechnic Institute
Quiz (10 point)

• (2 pts) What are “adjacency pair” and “backchannel”?

• (2 pts) When should the connection events better be suppressed?

• (4 pts) Explain the differences between multi-situated modules and multi-experts

• (2 pts) In saliency map, what region may be highlighted (or with higher temperature)?
  • Hint: list at least three factors
Human Social Learning
Overview

• How to scaffold learning process (cont)
  • Dynamic scaffolding
  • Externalize and modeling meta-cognition

• How to demonstrate to robot learner
  • Correspondence problem
  • Learning by doing, observation and critique
How to Scaffold the learning process?
Human teacher scaffold robot learner

• Human can help robot with hard problems in learning

• Three methods for scaffolding
  • Direct robot attention
  • Dynamic scaffolding
  • Externalize and modeling meta-cognition
Saliency map

Input video

Linear filtering

Image pyramids

Colors

Intensity

Orientations

Flicker

Motions

Center-surround differences and normalization

Feature maps

(12 maps)

(6 maps)

(24 maps)

(6 maps)

(24 maps)

Across-scale combinations and normalization

Conspicuity maps

Linear combinations

Saliency map

Winner-take-all

Attended location

camera focusing on parent

partner (infant or adult)
camera focusing on partner

Demonstrate to robot simulation

• Program robot to look at the most salient part
Dynamic scaffolding

• Adult (teacher) creates a learning situation of right level of complexity for the learner

• Mind the gap
  • What the learner has already mastered
  • What the learner can achieve with the aid of teacher

• Teacher’s role
  • Provide safety and intermediate attainable goals
Apply to LfD – Environmental scaffolding

• Context – Learning policy from demonstration

• Construct a trained hierarchy that include
  • Primitives
  • Sequence of primitives
  • Learned goal-directed tasks
  • Behaviors
Inspired by the hierarchy of animal behavior

Motion Primitives

Motion Sequence

Behavior

Goal-directed Tasks
Robot’s knowledge hierarchy
Pre-defined Primitives

- Basic actions available to robot
- Human has **no access to the internal states** of the robot
- Good or bad?

<table>
<thead>
<tr>
<th>Primitive</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pan Left</td>
<td>Pan Left 5° or continuously</td>
</tr>
<tr>
<td>Pan Right</td>
<td>Pan Right 5° or continuously</td>
</tr>
<tr>
<td>Tilt Up</td>
<td>Tilt Up 5° or continuously</td>
</tr>
<tr>
<td>Tilt Down</td>
<td>Tilt Down 5° or continuously</td>
</tr>
<tr>
<td>Move Forwards</td>
<td>Move forwards 10cm or continuously</td>
</tr>
<tr>
<td>Move Backwards</td>
<td>Move backwards 10cm or continuously</td>
</tr>
<tr>
<td>Turn Right</td>
<td>Rotate right by 5° or continuously</td>
</tr>
<tr>
<td>Turn Left</td>
<td>Rotate left by 5° or continuously</td>
</tr>
<tr>
<td>Increase Joint Angle(n)</td>
<td>Increase one of the six joint angle by 5°</td>
</tr>
<tr>
<td>Decrease Joint Angle(n)</td>
<td>Decrease one of the six joint angle by 5°</td>
</tr>
<tr>
<td>Move Arm (angle vector)</td>
<td>Move the arm to a given position using the six angles (angle vector)</td>
</tr>
</tbody>
</table>
Learning a new task

- Robot has learning and execution modes

- At learning mode, robot can be taught new competence
  - Sequences
  - Goal-directed tasks tasks
  - Behaviors
Learning sequence

• Sequence of pre-defined primitives

• Sequences are recorded without refer to internal/external states

• Sequences are fixed action patterns
  • E.g. move arm to a position = “readyArm”
Learning goal-directed task

• Learning action with external/internal states

• Trainer can select primitives, sequence and other goal-directed tasks

• Trainer informs the robot when task is completed
  • Pair goal condition with system states
Behavior

• Construct behavior based on tasks, sequences and primitives

• Similar to learning task, but no goal state is required

• Behavior can run continually based on current environment states

• Behavior can be built upon behavior
  • How?
Benefits of scaffolding?

- Build robot competencies by constructing **hierarchical** state/action maps
- Human can **teach** a robot new skills and **modify** skills that the robot may possess
- Robot can notify human **what it already has in skill sets**, and focus its sensor resources to the **relevant features of the skill being demonstrated**
Externalizing and modeling meta-cognition
Externalizing and modeling meta-cognition

"Put it in the green bin."
Simon: "OK"
• Read Section 2.4 in “Robot learning from human teachers”
  • Including papers in reference
  • Choose a reference paper you like and read thoroughly
  • Write a paragraph of review
  • Prepare 4-6 presentation slides on this paper
Reference


Demonstrate to Robot Learner
Overview

- Correspondence Problem
- Learning by doing, observation, and critique
Correspondence Problem

• LfD data set = (Mostly) state-action pairs

• Correspondence
  • How demonstrations are recorded?
  • What platform is used for execution?

• Correspondence problem arises from the difference of
  • Sensing capability
  • Physical embodiment
Mapping from Teacher to Learner

• Direct mapping
  • No correspondence problem
  • Demonstration is recorded in robot’s sensing states

• Not available option to all systems
  • Complex, coordinated motion on high degree of freedom
  • Controlling the robot physically may not be natural
Learning by doing

• Teleoperation
  • Input devices: Joystick, teleoperation suit, ...

• Kinesthetic teaching
  • Human teacher moves robot’s passive joint
  • Robot records trajectory in its own state space
  • Essentially a variant of teleoperation
Kinesthetic teaching using force control
Demonstration via Teleoperation
Learning by observation

- Human teachers demonstrates on their body
- Robot learner needs to accurately sense the teacher’s states and actions
First-person Teleoperation of humanoid robot
Teleoperation via speech

Adjust posture (center of mass)
Learning from Critique (Shaping)

- Robot practices the task
  - Select actions through exploration
- Teacher provides feedback to indicate the desirability of the exhibited behavior
Teaching by feedback

- Robot displays its learned behavior while being taught by feedback

- Benefits?
  - Increase responsiveness to teaching
  - Ensure teaching is focused on robot’s behavior based on its own policy
  - Feedback interface can be independent of the task domain
  - Require less expertise and cognitive load from trainer
Reward shaping

• How to use Reward signals?
  • Initially, reinforce tendency to correct behavior
  • Gradually, reward more difficult elements of the task → shaping

• Significantly more efficiently compare to feedback-only approach
Occasional Bias

- Instead of directly controlling all the agent’s actions
  - Human teacher occasionally bias the action selection

- Advantages?
  - Human doesn’t have to know all about how to perform the task
  - No need for undivided attention from human teacher
Assignment 11 (Due Nov 13)

- Next week, back to algorithms
- Read Chapter 4 “Learning Low-level Motion Trajectories”