Learning by Demonstration
A case study of learning Slide Around

Tuan Do
Brandeis University

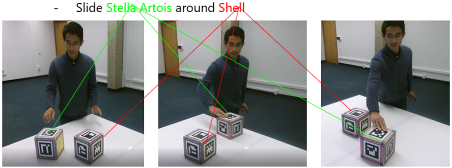
June 6, 2017

Introduction

- Learning from Demonstration or Imitation Learning (LfD) studies how computational and robotic agents can learn to perform complex tasks by observing humans.
    - Teach agents the concept of an action so that they can re-demonstrate it in a new context.
    - Adaptability to new situations is vital to next generations of AIs.
    - Learning from Demonstration leads to Learning from Interaction, i.e., learning new concepts is taken incrementally.
- Learning of action skills:
    - As difficult as that of composite shapes because of their *temporal-spatial* dynamics.
    - Multiple-step actions allows us to investigate interactive feedback.

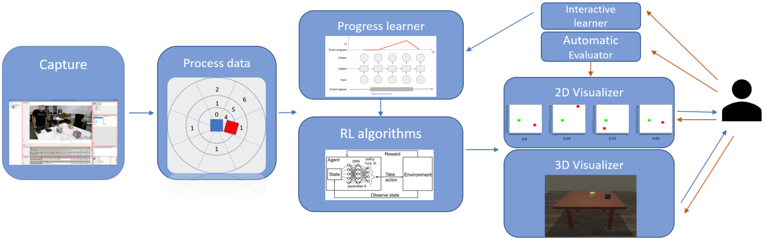
Introduction

- Case study: Learning of action skills: Slide Around
    - Making a rounding movement is a difficult concept that is only developed in later state of children.
    -
- Some simpler action skills are also tried, but none has the same difficulty as Slide Around: Slide Closer, Slide Next To, Slide Away, and Slide Past.



Framework

- Capture and annotation: an event capture annotation tool (ECAT) that captures human interaction with objects using Kinect sensors.
- Feature extraction and feature representation by Qualitative Spatial Reasoning (QSR), that enforces human bias on machine learning models.



Framework (cont.)

- Learning of a progress function to guide movement of actions:
  - Uses supervised learning with a recurrent neural network, in particular, LSTM.
  - Is similar to Inverse RL in spirit, i.e., to extract a reward signal from experts' behavior for reinforcement learning.
  - Produces a value from 0 to 1.
- Action reenacting algorithms: search and policy gradient algorithms in reinforcement learning setup.
  - Continuous space: action policy is modeled as Gaussian function of state.
  - Discrete space: the search space is divided into regions, and actions are moves between adjacent regions.
- Evaluation is based on human judgment on demonstrations of actions on 2-D and 3-D visualizers.

Experimental Design

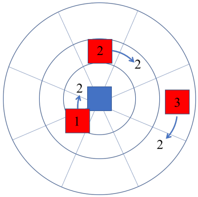
- Captures of 20 demonstrations of clockwise movement, 20 demonstrations of counter-clockwise movement, 2 performers for variation. Captures are taken using CwC Apparatus, and recorded by ECAT.
- Annotation is semi-automatically: performers and objects are tracked; action spans (beginning and end frames) and a description of the action are annotated in a separate session.
- Captures are sliced into equal-length chunks to be fed into LSTM to train the progress learner.
- Experiments with some algorithms: greedy search, beam search, REINFORCE, ACTOR-CRITIC.

Evaluation

- Human evaluation: Some hired annotators are shown demonstrations on 2-D and 3-D simulators, and asked to give a grade between 0 and 10.
- Automatic evaluation: Calculate the covering angle of the moving object around the static object; for one step, is positive if counter-clockwise, negative if clockwise.
  - Covering angle is absolute of the sum of this value for each step.
  - Score will be either 0, 1, or 0.5, based on two threshold values.
  - By calculate correlation with human judgment, the best value combination of thresholds are $alpha_1 = 1.1$ and $alpha_2 = 1.7$.

Results

- ACTOR-CRITIC gave a nice and interpretable solution: "Always go to the right", i.e., clockwise movement.
- Planning for a new demonstration can be facilitated by greedy selection on the action policy.
- Models run on discretized space work, but on continuous space do not work as well.
- Evaluation score:
  - Human evaluation: 6.7/10
  - Automatic evaluation: 0.7/1



Learning from Interaction (LfI) (ongoing)

- Use of simple communicative means (positive-ack and negative ack, and deixis/pointing) to improve learned action models.
    - Online corrections in the middle of a demonstration by clicking in a 2-D simulator to specify more proper locations after seeing agents demonstrate an action
    - Clicking in the 2-D simulator can be mapped to pointing in the 3-D communicative platform.
- Next step: Currently, we need some demonstrations to bootstrap the model; ongoing work on exclusively interactive learning.
- Next step: While directing Diana to pick up and move objects is sometimes clumsy, direct interaction with objects (by hand motion/gesture) can be used for demonstration.

Future Directions

- Path- versus manner- aspect of actions: for learners to make the distinction in performing "rolling a bottle" and "sliding a bottle".
- Symbolic versus feature-based learning: We should only consider LfD as one modality for teaching action skills to AI agents. A holistic approach to action learning is to teach trajectory action skills from demonstrations, and teach more complex skills by communicative means.
- Virtual reality LfD: Virtual reality (VR) can be used as a shared collaborative environment between humans and machines. Imagine a teaching classroom in a VR environment, and the lesson is the usage of a cutting tool, e.g., a knife, on some objects.