Mapping Instructions and Visual Observations to Actions with Reinforcement Learning

Dipendra Misra, John Langford† and Yoav Artzi*  
Cornell University, †Microsoft Research

Problem Statement

**Goal:** Map instructions to actions

Agent observes raw images and text to generate actions

Common approach:
- Decompose problem into different modules
- Design intermediate representations

Our approach: single learned model
- No intermediate representation required
- No need to build and train separate models

Blocks World

- Bisk et al. 2016
- 20 blocks, 81 actions each step
- Data: instructions and demonstrations
- Error is the sum of block distances between goal and final states

Observations

Processing

Instruction Text

CNN

LSTM RNN

Previous action

Images of current and previous states

Task specific

Softmax layers

Process repeats until STOP action

Few-Sample Reinforcement Learning for Natural Language Instructions

Learning approach: RL with task-completion problem reward

**Challenges:**

1. **Generalize to new tasks**
   - RL models generally trained and tested on one task

2. **Long action sequences**
   - Difficult to train with sparse problem reward $R_p$
   - Move Toyota to the immediate right of SRI, evenly aligned and slightly separated

3. **Limited language data**
   - RL is data hungry, often trained with a lot of data

**Example-specific reward:**
- Define a reward function for each example
- Optimize sum of RL objectives for each example

**Augment reward function with two shaping terms:**
- $F_1$ Encourage moving closer to the goal state
- $F_2$ Encourage following the demonstration

**Contextual bandit setting:**
- Maximize immediate reward
- Lower sample complexity

Results and Analysis

**Mean Execution Error**

- Stop
- Supervised
- REINFORCE
- DQN
- Our Approach

- 24/39% error reduction from supervised/DQN
- Agent often fails to stop or takes too many steps

- Closest agent got to goal during execution

**Varying Amount of Supervision**

- Mean execution error, development results

- Some demonstrations necessary, but can do well with relatively little

**Unknown Words and Planning**

- Development Error
- No Unknown Words
- 1 Unknown Word
- 2 Unknown Words
- With Planner

- Sensitive to unknown words, planning still key problem