Lecture 11

Artificial Neural Networks (1 of 4)

Friday, February 11, 2000

Kiran Nandivada
Department of Computing and Information Sciences, KSU

Readings:
"Incorporating Advice into Agents that Learn from Reinforcements"
Richard Maclin and Jude W. Shavlik

Presentation Overview

Paper
- "Incorporating Advice into Agents that Learn from Reinforcements"
  - Authors: Richard Maclin and Jude W. Shavlik, Computer Sciences Department, University of Wisconsin

Overview
- Learning from reinforcements by accepting advice from an external observer
  - The system accepts the advice
  - The external observer can provide advice at any time

Presentation Outline

Issues
- Is the advice given by the external observer used effectively?
- Does it matter in this type of learning when the advice is given?
- Key strengths - the use of external observer enhanced the learning process
- Key weaknesses - accepts only single advice at a time

Outline
- Advice taking
  - Proposed a strategy where several steps described by Hayes-Roth, Klahr, and Mostow (1981), can be achieved using reinforcement learning.
- Experiments
- Test Environment
- Results
- Future work
- Summary

Terminology

Reinforcement learning
- Reward or Reinforcement
  - Feedback provided to the agent for the action it performed in the previous state
  - Task of learning
    - The agent learns from this reward and chooses actions that produce highest cumulative reward (Mitchell, Ch. 13)

Q-learning
- The agent learns a numerical evaluation function defined over states of actions, and then implement an optimal policy in terms of this evaluation function (Mitchell, Ch. 13)

Connectionist Q-learning
- The utility function is implemented as neural network, whose inputs describe the current state and whose outputs are the utility of each action

Advice-taking

Step 1 - Provide advice to the agent
  - Advice is provided by the external observer whenever the observer feels appropriate

Step 2 - Convert the advice to an internal representation
  - Expression of advice is in the form of a simple programming language and list of terms which specific certain tasks.

Step 3 - Convert the advice into an usable form
  - Requires a compiler for certain task specific terms

Step 4 - Integrate the reformulated advice into the agent's current knowledge base
  - Used an extended KBANN approach

Example - Agent learning to play a video game
  - A sample version of the advice provided to the agent
Advice-taking

- A sample version of advice
- Advice

IF An Enemy IS (Near West) THEN
  An Obstacle IS (Near North)
  MULTIATION
  MoveEast MoveNorth
END;

WHEN Surrounded
  OKtoPushEast
  An Enemy IS Near
REPEAT
  PushEast
  MoveEast
UNTIL
  ¬ OKtoPushEast
  ¬ Surrounded

Advice-taking

- Network showing the advice added by adding hidden units that correspond to the advice

Advice-taking

- Allows advice that contains multi-step plans

Advice-taking

- Allows advice that refers to previously defined terms

Experimentation

- Step 5 - Judging the value of advice
  - Introduces a Q-learning concept to “wash-out” a poor advice
  - Empirically evaluate the new advice
  - Retracts or counteract a bad advice

- Experiments
  - Goal - Empirically evaluate whether this particular approach of providing advice is better
  - Hypothesis 1 - System takes advantage of the advice
  - Hypothesis 2 - Observer provides appropriate advice to the agent at any time during the training

- Test Environment
  - Agent performs certain actions which include moving and pushing in the directions East, West, North, and South and doing nothing
**Experimentation**

- Assumes an agent-centered model partitioned into sectors
- Agent calculates the percentage (input to the network) of the type of the object occupied in each sector

**Methodology**

- Train the agents for a fixed number of episodes
- Choose an initial topology
- Provide advice to the agent
- Four forms of advice are provided to the agent (ElimEnemies, Surrounded, SimpleMoves, NonLocalMoves)

**Result**

- Experiment 1
  - Train the agent initially
  - Measure the value of adding advice
  - Add the advice and measure the test set reinforcement

**Results**

<table>
<thead>
<tr>
<th>Advice added</th>
<th>Enemies</th>
<th>Rewards</th>
<th>Survival time</th>
</tr>
</thead>
<tbody>
<tr>
<td>None</td>
<td>0.15</td>
<td>3.09</td>
<td>32.1</td>
</tr>
<tr>
<td>SimpleMoves</td>
<td>0.28</td>
<td>3.79</td>
<td>39.6</td>
</tr>
<tr>
<td>NonLocalMoves</td>
<td>0.26</td>
<td>3.95</td>
<td>39.1</td>
</tr>
<tr>
<td>ElimEnemies</td>
<td>0.44</td>
<td>3.50</td>
<td>38.3</td>
</tr>
<tr>
<td>Surrounded</td>
<td>0.30</td>
<td>3.48</td>
<td>46.2</td>
</tr>
</tbody>
</table>

Exp 1 - Average test performance on the four tasks in paper (significant with 99% confidence)

**Significance**

- Reported the gains obtained over the baseline or without the addition of advice
- Initial training was for 1000 episodes and system training after adding advice was for 2000 episodes and the baseline for 3000 episodes

**Experiment 2**

- Insert advice at different times during the training (0, 1000, and 2000 episodes)
- Convergence to same amount of reinforcement irrespective of the time the advice was provided
- Observe whether each task corresponding to each piece of advice is performed

**Summary**

- **Future Work**
  - Accepting multiple pieces of advice at different times during training
  - Evaluate periodic retraining or “replay” on certain pairs of states
- **Possible improvements**
  - a) Temporal difference method
  - b) Also, can improve on their advice taking strategy
  - c) Use EBL to improve on pieces of advice and use it to solve any advice-taking problems
- **Content Critique**
  - Key Contribution - A good example of how the learner can accept “general advice” at any time during the training (agent learning naturally)
  - Strengths - Proved that the advice improves the expected rewards
  - Weaknesses - No reference to convergence in terms of error and generalization
- **Presentation Critique**
  - Audience - AI, Robotics (principally concerns game playing strategically)
  - Positive points - Good introduction and explanation regarding advice taking steps
  - Negative points - Results - Did not mention what kind of statistical methods were used to get the significance they reported.