ECE 8843 Fall Semester, 2012

HOMEWORK 3 – LEARNING ROBOT CONTROL STRATEGIES

Assigned: Oct. 19, 2012 Due: Nov. 2, 2011 – 12:05pm

Robby, the Soda-Can-Collecting Robot, has the job of cleaning up his environment by collecting empty soda cans. Robby's world consists of 100 site locations laid out in a 10 x 10 grid. Various sites have been littered with soda cans (but with no more than one can per site). Robby isn't too intelligent, and his eyesight isn't that great. From wherever he currently is, he can see the contents of one adjacent site in the north, south, east, and west directions, as well as the contents of the site he's currently in. A site can be empty, contain a can, or be a wall (which surrounds Robby's world).

For each cleaning session, Robby can perform exactly 200 total actions. Each action consists of one of the following seven choices: move north, move south, move east, move west, move in random direction, stay put, or pick up can. Each action generates a penalty of 1 point. In addition, if Robby is in the same site as a can and picks it up, he gets a reward of 10 points. However, if he bends down to pick up a can at a site where there is no can, he is fined 1 point. If he crashes into a wall, he is fined 5 points and bounces back into the current site. Clearly, Robby's reward is maximized when he picks up as many cans as possible, without crashing into any walls or bending down to pick up a can when no can is there. Your assignment is to write code for a learning algorithm (reinforcement learning, neural networks, or genetic algorithms) that learns an optimal control strategy for Robby the Robot.

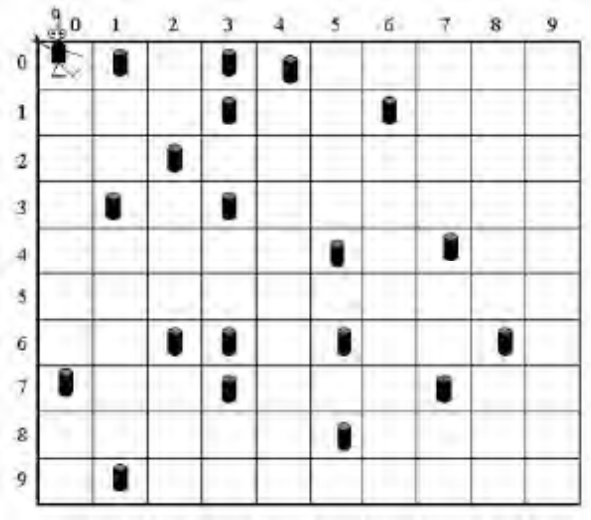
A. SETTING UP EVERYTHING. The world map "Can.wld" in which Robby will navigate is defined by the following parameters:

Origin x_position y_position
Defines the starting (x,y) position of the robot

Can x_position y_position
Defines the (x,y) location of a can

.

```
Origin 0 0
Can 1 0
Can 3 0
Can 4 0
Can 3 1
Can 6 1
Can 2 2
…
Can 1 9
```



B. ENCODING OF STRATEGIES

For every possible state, Robby's learned strategy should give the action he should take when in that state. A state consists of the five different sites (north, south, east, west, current), each with three possible types of contents (wall, empty, can). Thus there are $3^5 = 243$ possible situations in which the robot can be in (and which need an action defined). Given this construct, a part of a learned strategy is as shown:

Situation					Action
North	South	East	West	Current Site	
Empty	Empty	Empty	Empty	Empty	MoveNorth
Empty	Empty	Empty	Empty	Can	PickUpCan
Empty	Empty	Empty	Empty	Wall	MoveRandom
Empty	Empty	Empty	Can	Empty	MoveWest
⋮	⋮	⋮	⋮	⋮	
Wall	Empty	Can	Wall	Empty	MoveEast
⋮	⋮	⋮	⋮	⋮	
Wall	Wall	Wall	Wall	Wall	StayPut

For example, Robby's situation in the above example is:

North	*South*	*East*	*West*	*Current Site*
Wall	Empty	Can	Wall	Empty

After learning a strategy, to decide what to do next, Roby could simply look up his situation in a strategy table and find the corresponding action to take. Thus, the strategy employed for this situation would be: *MoveEast*

C. AN EXAMPLE SOLUTION – SETTING UP THE PROBLEM

Imagine you decide to utilize genetic algorithms for your learning algorithm. In this case, the "chromosome" to be evolved by the GA could just be a listing of the 243 actions (where each gene represents a situation). If the actions are numbered as:

0: move north
1: move south
2: move east
3: move west
4: move in a random direction
5: stay put
6: pick up can

Then the chromosome representing the strategy in the example above would be: 0643…2…5

D. TEST COMPLETE SYSTEM. Test your learning algorithm using the world above. Run your code 5 times (with Robby starting at different starting locations) and calculate your score after 200 actions.

E. WRITE UP THE FOLLOWING (submit as a single pdf file called *yourlastname*.pdf): 1) A description of your learning algorithm, including its parameter values (for example: GA population size) [description should be no more than half-page in length], 2) The set of actions learned by your algorithm based on the world above (i.e. set of actions associated with each site on the map), 3) The performance score for your five runs and the starting locations utilized, and 4) A list of the sequence of actions (for one run) that illustrate your program's ability for Robby to clean its environment (remember to provide the starting location).

F. SUBMISSION. For homework credit, turn in your code, a README documenting how to run your code, and your .pdf writeup.