note 160 views

Open Source Assignment 4 (Updates)

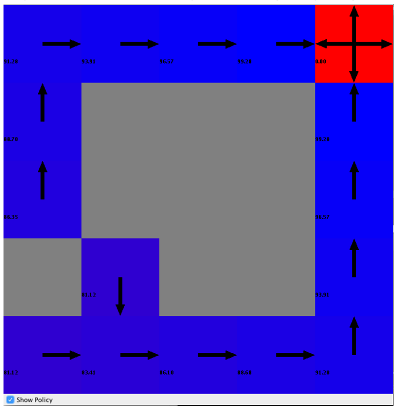
Hi All,

Considering the number of students that have cloned the open source assignment 4 repository, I thought that I should mention that there have been a significant number of bug fixes and improvements since its inception. If you downloaded a copy of the code early on, I highly advise that you pull/clone the latest updates.

The GitHub Repository: https://github.com/juanjose49/assignment4
The open source Assignment 4: https://github.com/juanjose49/assignment4/tree/master/src/burlap/assignment4

Thanks,
Juan

UPDATE:
I just merged some code changes submitted by Jim Gorman/David Musielewicz that generate really nice policy maps using BURLAP code. If you pull the most recent code changes you're now going to see three new variables at the top of the two launchers: showValueIterationPolicyMap, showPolicyIterationPolicyMap, and showQLearningPolicyMap. Setting any of them to true will generate a policy map as seen below:



Notice that there's a checkbox at the bottom of the window to show the actual arrows for the policy. Also, consider only setting one of those variables to true since the policy maps aren't identified by algorithm and it may be difficult to discern which policy map belongs to which algorithm. Thanks again guys!

For anyone seeing this for the first time, this is the sort of output that you'll get out of the box by simply executing the main classes Easy/HardGridWorldLaunchers as Java Applications:

/////Easy Grid World Analysis/////

```
This is your grid world:
[0,0,0,0,0]
[0,1,1,1,0]
[0,1,1,1,0]
[0,1,1,0,0]
[0,0,0,1,0]

//Value Iteration Analysis//
Passes: 1
...
Passes: 15
Value Iteration,175,157,9,13,17,12,10,10,12,11,11,9,14,10,9
This is your optimal policy:
[>,>,>,>,>]
[^,*,*,*,^]
[^,*,*,*,^]
[^,*,*,>,^]
[^,<,<,*,^]

//Policy Iteration Analysis//
Total policy iterations: 1
...
Total policy iterations: 15
Policy Iteration,67476,2464,556,10,12,15,11,10,12,10,9,11,13,17,12
Passes: 13
This is your optimal policy:
[>,>,>,>,>]
[^,*,*,*,^]
[^,*,*,*,^]
[^,*,*,^,<]
[^,<,<,*,>]

//Q Learning Analysis//
Q Learning,22,84,28,56,10,12,21,11,18,39,12,10,10,11,18
Passes: 13
This is your optimal policy:
[>,>,v,>,<]
[^,*,*,*,v]
[^,*,*,*,^]
[^,*,*,>,v]
[^,<,<,*,v]
```

```
//Aggregate Analysis//
The data below shows the number of steps/actions the agent required to reach
the terminal state given the number of iterations the algorithm was run.
Iterations,1,2,3,4,5,6,7,8,9,10,11,12,13,14,15
Value Iteration,175,157,9,13,17,12,10,10,12,11,11,9,14,10,9
Policy Iteration,67476,2464,556,10,12,15,11,10,12,10,9,11,13,17,12
Q Learning,22,84,28,56,10,12,21,11,18,39,12,10,10,11,18
The data below shows the number of milliseconds the algorithm required to generate
the optimal policy given the number of iterations the algorithm was run.
Iterations,1,2,3,4,5,6,7,8,9,10,11,12,13,14,15
Value Iteration,49,4,3,5,6,8,8,19,6,7,8,6,2,2,2
Policy Iteration,5,0,0,1,1,1,1,2,2,3,3,3,3,3,3
Q Learning,153,3,4,5,3,4,4,4,3,3,2,3,3,3,4
```

hw4

Updated 7 days ago by Juan J. San Emeterio

followup discussions *for lingering questions and comments*

● Resolved ○ Unresolved

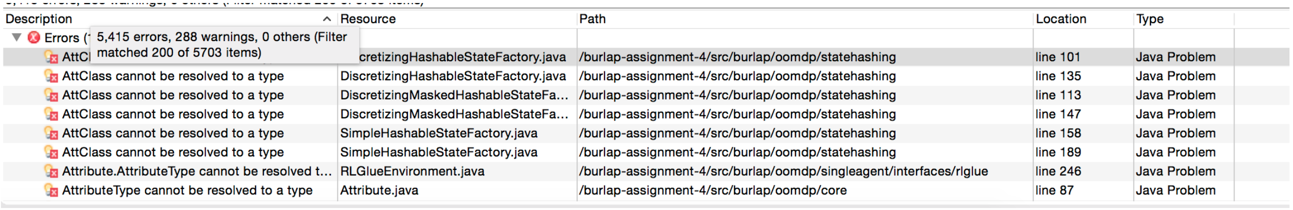
Lan Wu 10 days ago
Hi Juan - This is amazing, thank you. I installed ant, JDK, ran ant successfully. However now when i try to run any of the three, a ton of errors appear, mostly like "cannot find symbl". I'm a complete java novice -- but assuming these means that there is some definition file that should be somewhere? Any help would be greatly appreciated!
Thanks!!

Lan Wu 10 days ago What command should i be using to run the three files - do i need to compile with 'javac', and then run 'exec'? And should i be in assignment5-master, or in \src\burlap\assignment4?

Juan J. San Emeterio 10 days ago My recommendation is that you download the repository and import it using Eclipse as a project. You can then navigate to the files and simply run it as an application.

Kevin Davenport 10 days ago After building for ant is it possible to run directly such as $ java -cp burlap.jar ../build.burlap.assignment4.EasyGridWorldLauncher

GAURAV PURI 8 days ago



Description	Resource	Path	Location	Type
Errors (5,415 errors, 288 warnings, 0 others (Filter matched 200 of 5703 items)				
AttCl	cretizingHashableStateFactory.java	/burlap-assignment-4/src/burlap/oomdp/statehashing	line 101	Java Problem
AttClass cannot be resolved to a type	DiscretizingHashableStateFactory.java	/burlap-assignment-4/src/burlap/oomdp/statehashing	line 135	Java Problem
AttClass cannot be resolved to a type	DiscretizingMaskedHashableStateFa...	/burlap-assignment-4/src/burlap/oomdp/statehashing	line 113	Java Problem
AttClass cannot be resolved to a type	DiscretizingMaskedHashableStateFa...	/burlap-assignment-4/src/burlap/oomdp/statehashing	line 147	Java Problem
AttClass cannot be resolved to a type	SimpleHashableStateFactory.java	/burlap-assignment-4/src/burlap/oomdp/statehashing	line 158	Java Problem
AttClass cannot be resolved to a type	SimpleHashableStateFactory.java	/burlap-assignment-4/src/burlap/oomdp/statehashing	line 189	Java Problem
Attribute.AttributeType cannot be resolved t...	RLGlueEnvironment.java	/burlap-assignment-4/src/burlap/oomdp/singleagent/interfaces/rlglue	line 246	Java Problem
AttributeType cannot be resolved to a type	Attribute.java	/burlap-assignment-4/src/burlap/oomdp/core	line 87	Java Problem

I tried to import the project into GIT using Eclipse . On importing i get 5415 errors. (I have no prior experience with JAVA / eclipse)

Also , does anyone know how to run these 3 files using CLI ?

GAURAV PURI 8 days ago Seems like it was issue with Compiler version.

Kevin Davenport 8 days ago If you're on mac use homebrew to install the latest java then grab eclipse from the official site.

● Resolved ○ Unresolved

James Kirk Muse 10 days ago

In your code, what is the difference between iterations and intervals? It looks like it is just used to create an incremental value.
int increment = MAX_ITERATIONS/NUM_INTERVALS;

Also, what reward function is being used? I see the code below, but is this code subtracting points for each step that is not the goal state?

RewardFunction rf = **new** BasicRewardFunction(maxX,maxY); //Goal is at the top right grid

Juan J. San Emeterio 10 days ago MAX_ITERATIONS is the number of iterations you would like the last experiment to run.
NUM_INTERVALS is the number of experiments you would like to run.
Ultimately the two are used to calculate the number of iterations to run per experiment and provide an upper bound on how many experiments you would like to run.

If you go to the class you'll find that for every state, except the goal state, return -1 and 100 if it's the goal state.

James Kirk Muse 7 days ago Thanks Juan!

● Resolved ○ Unresolved

Kevin Davenport 10 days ago
Juan for President!

Juan J. San Emeterio 10 days ago Thanks! I'm glad you found the code helpful.

Juan

Kevin Davenport 10 days ago Interesting just noticed you can pass directions to the world for interaction!

Juan J. San Emeterio 9 days ago Yeah, there are also ways to feed the episodes (or simulations) that are calculated using the policies into the GUI. This allows you to see step-by-step the agent traversing the grid world. Ultimately I didn't use any of that functionality because it seemed useless for our purposes. BURLAP is a pretty impressive piece of software given the time to learn it properly.

Juan

● Resolved ○ Unresolved

Pauline Chow 10 days ago
Thanks for sharing your code. Each assignment has a guardian angel(s)...so it played out!

Juan J. San Emeterio 9 days ago Glad i could help!

● Resolved ○ Unresolved

Xinqiong Yu 10 days ago
Juan, Thanks for sharing! When I run your "EasyGridWorldLauncher", a GUI appears, I can put in "east" "west" in the console for it to execute, but how to run the learning algorithm to auto generate the result you posted? Thanks!

Juan J. San Emeterio 10 days ago By default, the outputs should be printed automatically to the console. Only thing to notice is that if you close the window with the GUI you will also terminate the execution and the app will not print the expected output.

Juan J. San Emeterio 10 days ago Also notice the Boolean variables at the top of each launcher. Those booleans dictate what gets runs. You can disable the visualized functionality with those booleans if it's getting annoying to have to close it after each execution.

● Resolved ○ Unresolved

Lan Wu 10 days ago

● Resolved ○ Unresolved

Anshul 10 days ago
Hey Juan,

Thanks for all the code really saved a lot of time.

Can you please help me get to the code that controls the stochasticity of the agent?

I tried changing the directionProbs here but not sure if this was it as the agent got stuck(could be my grid) but i just want to make sure i am messing with the right area of the code.

```
for(int i = 0; i < 4; i++){
                        if(i == direction){
                                directionProbs[i] = 0.8;
                        }
                        else{
                                directionProbs[i] = 0.2/3.;
                        }
                }
```

I would appreciate the help

Juan J. San Emeterio 9 days ago Anshul, you do seem to be in the right place but why it isn't working, I'm not sure. There are still a few parts of this code that are a mystery to me and how the reward function is being exactly used by the model is one of those parts. If you figure out how to make this work (via tweak or code change) please feel free to submit a pull request so that others can benefit.

Juan

● Resolved ○ Unresolved

Brent Wagenseller 9 days ago
Hey Juan,

Fantastic stuff!

Have you seen the section of the BURLAP tutorial that plots policy? It seems pretty awesome. If you havent seen it, you may want to give it a look!

http://burlap.cs.brown.edu/tutorials/bpl/p5.html

David Musielewicz 9 days ago Brent,
I've been playing with this trying to get it to work with Juan's code but haven't had any luck yet. Were you able to use it to generate the graphs?

Brent Wagenseller 9 days ago I haven't just yet.

Juan J. San Emeterio 9 days ago That's a pretty cool find. If either of you get it up and working please submit a pull request for other people's benefit. I've gotten around the ugly policy maps by using the PowerPoint I uploaded here: @758.

Juan

David Musielewicz 8 days ago Alright, I got it working. I altered a lot of the code for my own readability so it'd probably be easier for you to just add the below function to your AnalysisRunner.java file.:

```
public void simpleValueFunctionVis(ValueFunction vf, Policy p, State initialS, Domain domain){
        List<State> allStates = StateReachability.getReachableStates(
                initialS,
                (SADomain)domain, new SimpleHashableStateFactory());
        ValueFunctionVisualizerGUI gui = GridWorldDomain.getGridWorldValueFunctionVisualization(allStates, vf, p);
        gui.initGUI();
}
```

Then to use the function add the below line after the 'System.out.println("\n\n");' in your value and policy iterations.

```
runValueIteration:
        simpleValueFunctionVis((ValueFunction)vi, p, initialState, domain);
runPolicyIteration:
        simpleValueFunctionVis((ValueFunction)pi, p, initialState, domain);
runQLearning:
        simpleValueFunctionVis((ValueFunction)agent, p, initialState, domain);
```

Finally, don't forget to check the "Show Policy" checkbox at the bottom to see the overlay of your policy. Let me know if this is too confusing, but I hope this helps!

Jim Gorman 7 days ago Juan and David: I just submitted a pull request for this.

Juan J. San Emeterio 7 days ago Merged!

Brent Wagenseller 7 days ago great work on the map, gentlemen!

● Resolved ○ Unresolved

Asela Wijeratne 9 days ago
Hey Juan,
This is fantastic and I am going to use this for the assignment as I am running out of time (spent too much time on mdptoolbox).
However, being not java savvy, I am still confused as how to proceed. I cloned your git repo and used 'ant' to build it. Is there a way to call jar files from the terminal?

Brent Wagenseller 9 days ago Are you using Eclipse? Eclipse is probably the most well-known IDE for Java. I am not quite sure Juan made this for ant (although I could be mistaken)

If you are, copy the entire 'assignment4' directory and place it in your Eclipse workspace directory. From there, open Eclipse, and import the project (located in workspace/assignment4).

If you are not using Eclipse....you may REALLY want to consider it. It only takes 15ish minutes to get acquainted with everything you need to know about it (mainly, how to load / open a new project and how to run a project).

Eclipse is at https://eclipse.org/downloads/ ; its free, and it has a GUI for Windows / Mac / Ubuntu (maybe even Red Hat as well?)

Eclipse lists all available objects once you type a dot after an object, and it displays autocorrections for the easier mistakes to solve (which is most of them); it even will recommend libraries to import if you referenced a library that is not imported. Its all that and a bag of chips.

Asela Wijeratne 9 days ago I think I got it to work and thanks a lot for the information. This saved me a lot of time.

I can enjoy Thanksgiving a little better now although still a long way to go!

Brent Wagenseller 9 days ago If you need any help, let us know here.

To be clear: I did not use Eclipse for project 2 (was a bit over my head, and I couldnt get it to quite work). Project 4 is geared for Eclipse, though - so take advantage! :)

Tyson Bailey 8 days ago I didnt know it was geared for eclipse, ive just been using ant.... lol

Pauline Chow 7 days ago i used both for assignment 3. But for some reason running in CLI for assignment 4 doesn't always work. Eclipse works for me in assignment 4 and I don't need to rebuild every time I change something.

● Resolved ○ Unresolved

Pauline Chow 8 days ago
Question about outputs related to plotting for the final write up. In @748 the "standard" plot for MDPs are cumulative rewards.

The open sourced code provided final outputs that provides the number of steps/actions the agent required to reach the terminal state. How do we add an output for cumulative rewards when running policy and value iterations? Is this not the expected output?

Before asking this question, I explored the burlap tutorial which provides the basic code that outputs cumulative steps and rewards for Q-learner using the "LearningAgentFactory". Unfortunately value and policy iterations are not learners and there is no plug and play for these algorithms. I see a way to solve this by adding functions in the **burlap.assignment4.util. AnalysisAggregator** but need some tips, as I am not as boss at Java or burlap as you all.

Thanks so much.

Kevin Davenport 8 days ago movement.java under assignment4.util gives probabilities to movement

Pauline Chow 7 days ago Are those value we want to map? We would get probabilities for each state / square in Grid World. Any thing else I am missing about this?

Kevin Davenport 7 days ago If you look in some of the supporting files like movement, runner, and rewards you'll see that all states are -1 which means the agent should want to reduce the steps necessary to get to the +100 terminal states, so I think this means 9 steps = (-9+100), 28 steps = (-28+100) and so on. This is my naive understanding at least :)

David Musielewicz 7 days ago Kevin this is spot on from what I'm seeing as well. If you debug the code and look at the 'ea' variable you'll see the rewards list which is all -1's until the final 100 is reached.

Howard Wayne Kim 6 days ago Episodeanalysis has a rewardSequence field that gives you give you the sequence of rewards for a given episode (and that episode follows a policy generated from value or policy iteration). Since you probably want an average over several episodes, keep running that policy, summing up the rewards, then take the average.

Here is my code for that that I put in AnalysisRunner.java from Juan's code:

```
ea = p.evaluateBehavior(initialState, rf, tf);
                        AnalysisAggregator.addStepsToFinishPolicyIteration(ea.numTimeSteps());
                        AnalysisAggregator.addIterationsPolicyIteration(pi.getTotalPolicyIterations());
                        List<Double> allRewards =  new ArrayList<Double>();
                        double sumRewards = 0;
                        for (int k = 0; k < 250; k++) {
                                ea = p.evaluateBehavior(initialState, rf, tf);
                                allRewards = ea.rewardSequence;
                                for (int j = 0; j< allRewards.size(); j++) {
                                        sumRewards += allRewards.get(j);
                                }
                        }
                        sumRewards = sumRewards / 250;
                        AnalysisAggregator.addRewardsPolicyIteration(sumRewards);
```

Also, does anyone know how to make this code block pretty and colorful?

● Resolved ○ Unresolved

Jim Gorman 8 days ago
Personally I prefer IntelliJ. If there's anyone else like me, here's what you need to do to use Juan's code with IntelliJ.

1. Clone it (or fork it).
2. Open the project folder in IntelliJ.
3. You will see a pop-up dialog that says 'Non-Managed pom.xml file found'. Select 'Add as Maven Project'.
4. Build > Rebuild Project.
5. You will see an error 'java: strings in switch are not supported in -source 1.6'.
6. Open pom.xml, and edit the following text in bold (from 1.6 to 1.8):

```
<groupId>org.apache.maven.plugins</groupId>
<artifactId>maven-compiler-plugin</artifactId>
<version>3.0</version>
<configuration>
  <source><strong>1.8</strong></source>
  <target><strong>1.8</strong></target>
</configuration>
```

Click 'Save'.
Project > Rebuild Project.

You will see a pop-up that says 'Maven projects need to be imported'. Click 'Enable Auto-Import'.

Click 'Save'.
Project > Rebuild Project.

You should be all set.

P.S. I sent Juan a pull request for this fix.

Juan J. San Emeterio 7 days ago Merged!

● Resolved ○ Unresolved

Anonymous 8 days ago

1. When I run Juan's code, I see lots of output.
2. I've also noticed that I must keep the GUI open to let Juan's code generate the full output (even though the GUI has nothing to do with the output).

My Question:

- Does the output from Juan's code give us everything we need for our analysis? (Save for the fact that it would be best to **plot** the output, etc.?)

Pauline Chow 7 days ago I am plotting the values from the output, steps to reach terminal state - both it's values and cumulative "steps" values (not sure if this 2nd part is right). My question above related to rewards is asked for any additional functions we can apply to get "reward" (or a similar measure) for value and policy iterations.

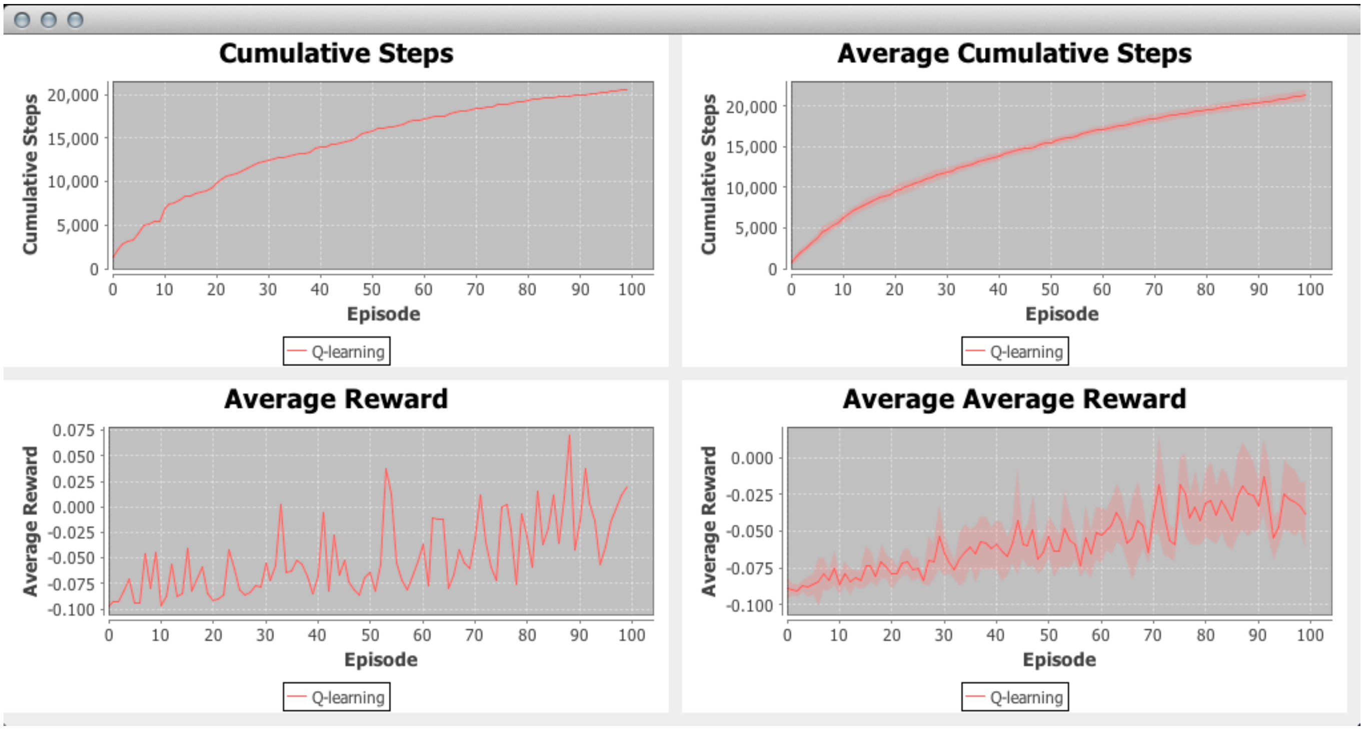
● Resolved ○ Unresolved

Anonymous 8 days ago
With BURLAP, we get the graphs below for free.

Has anyone figured out if it's worthwhile to include these too?

Should someone (perhaps me?) submit a pull request to add these to Juan's code?

Or are these already part of Juan's code and I just didn't notice them?



Gautam Salhotra 7 days ago Does this also work for planners like value iteration and policy iteration? LearningAgentFactory can only return learning agents with generateAgent(), not stochastic planners like VI or PI (I tried this).
The tutorial states

> In this section we will make use another EnvironmentObserver called PerformancePlotter to record a learning algorithm's performance and compare it to another learning algorithm.

Pauline Chow 7 days ago These graphs do not work for policy and value iterations since its a module/method for "LearningAgentFactory".

Jim Gorman 6 days ago I think Pauline is correct.

● Resolved ○ Unresolved

Kevin Davenport 7 days ago
I must be missing a fundamental concept. Would more attempts and time mean something liek Q-Learning "learns" and gets better thus needs less time to find the optimal policy? In other words shouldn't these lines be decreasing instead of increasing?

Juan J. San Emeterio 7 days ago Hey Kevin,

I'm seeing the same behavior as you and I think it is justified by the fact that the way Q Learning works is by running N number of iterations where the agent starts from the initial location and stumbles through the grid to the goal location. After each iteration, the sequence of actions that the agent took are hashed and recorded for later use when calculating a policy. The more iterations you run, the more data needs to be crunched when calculating an optimal policy.

By no means do I consider myself an expert on this matter so if anyone else can offer a better explanation please do so.

Thanks,
Juan

Kevin Davenport 7 days ago Thanks Juan, what you're saying makes total sense if you reconcile it with this: http://mnemstudio.org/path-finding-q-learning-tutorial.htm
I'm looking through the runner and movement files, but I fail to see where you constrained QL to always start in the same spot. Would it be more successful if we allowed it to start in a random place?

Juan J. San Emeterio 6 days ago The code that indicates that the algorithm will run from the initial state is a little bit hidden. Basically when the GridWorldLauncher runs the Q Learning algorithm it is also passing along the environment that was created in the first few lines of the main method. When the environment is created, it is initialized with the initial state. Then in the runQLearning method, after every episode, the environment is reset as seen below:

```
for (int i = 0; i < numIterations; i++) {
        ea = agent.runLearningEpisode(env);
        <strong>env.resetEnvironment();</strong>
}
```

In regards to you follow up questions about restarting from random places, I'm not really sure. The way I see it is that starting the algorithm from the same location each time allows you to get sequences of actions and rewards that can be compared directly to one another. By starting in random locations you will not be able to compare sequences directly and as a result will not be able to calculate the value of an action-state using the reward achieved at the end of a given sequence.

Juan

● Resolved ○ Unresolved

Jeffrey Tagen 7 days ago
Hi Juan,

I've been playing with your code for about a week, love what you've done, thanks for all the effort.

A quick question - I can't seem to find the source of the movement probabilities. Running through the interactive session for a map, it's clearly non-deterministic, but I can't for the life of me figure out where this is. Any advice?

Jeff

Jeffrey Tagen 7 days ago Silly me, assumed the Movement class was built into the burlap GridWorld examples, not your own. Found it :-)

● Resolved ○ Unresolved

Anonymous 7 days ago
Are the two examples EasyWorldGrid and HardWorldGrid sufficient for the assignment ? Or are we also looking at some other MDP Problem ?

Juan J. San Emeterio 7 days ago Hey Anonymous,

I asked a question very similar to yours a couple of days back and Prof. Isbell said that it was fine to use the same sort of problem with two different difficulties. As mentioned in the assignment, just make sure that the harder problem has a larger number of states than your easy problem.

edit: see @718

Juan

● Resolved ○ Unresolved

Daniel Tixier 7 days ago
Has anybody modified it to run the algorithms until they converge? Unless I'm missing something they are run for a set number of iterations, with no guarantee that what they output is correct. In fact, the algorithms often return different optimum policies.

Kevin Davenport 7 days ago I suppose if you know the optimal amount of actions to get to the end state you can see which output arrays contain that number first no? For the small 5x5 grid the optimal steps are 9.

Andy Tan 7 days ago The thing is for some problems (like in RL1, with a large, negative terminal state) the optimal policy actually goes around the negative terminal state and takes more steps. It won't be easy to always know what number of actions correspond to an optimal policy.

Kevin Davenport 7 days ago Hmm yeah that is a great point, not sure what to do in that case with this code.

Andy Tan 6 days ago I've modified the ValueIteration and PolicyIteration to run until the value of each cell converges to within 0.001 and am now seeing that ValueIteration and PolicyIteration converge to the same maps. For example with ValueIteration, I did this by changing the declaration of `vi` in runValueIteration of AnalysisRunner.java:

```
                    vi = new ValueIteration(
                                domain,
                                rf,
                                tf,
                                0.99,
                                hashingFactory,
                                /*changed this*/ 0.001, numIterations);
```

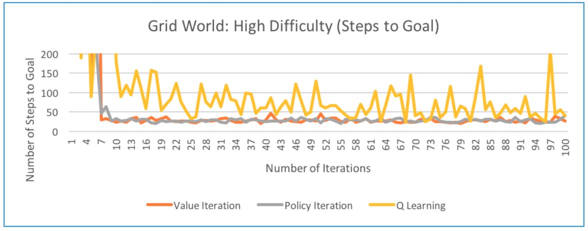
However, does anyone know how to terminate Q-learning when it "converges"? I'm not seeing a similar MaxDelta parameter.

Juan J. San Emeterio 6 days ago Hey Guys,

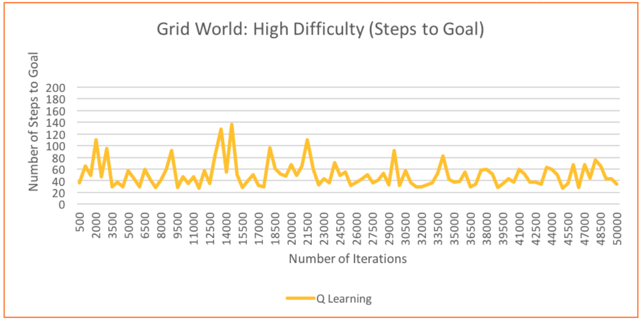
I think there's a bit of confusion... The whole reason I modified the input parameters for MaxDelta for the different algorithms to run until the max number of iterations were reached was so you could easily run a comparison between all three algorithms. Like you may have noticed, there is no MaxDelta parameter for Q Learning.

Convergence doesn't need to be measured by the MaxDelta, you can easily tell when the algorithms have converged by the fact that the number of steps required to reach the goal flattens out to a near constant number. This is your algorithm converging! What you'll notice if you plot the number of steps needed to reach the goal by the three algorithms, especially in the hard version of MDP, is that Value Iteration and Policy Iteration will converge almost immediately after they start running iterations. However, Q Learning will continue to have wild variability for a much longer time.

Take my graph below for instance. Value Iteration and Policy Iteration converged right around 7-10 iterations. Q Learning, on the other hand continues to suffer from variability all the way through the first 100 iterations. Clearly, Q Learning has not converged while the other two algorithms have.



That obviously leads to the next question: How long will it take for Q Learning to converge? In order to answer that question myself, I used an absurdly high max number of iterations and this is the graph that I obtained below. What you'll notice is that after 500 iterations, the number of steps never reach 200 steps like it did in the previous graph where the max number of iterations was only 100. Clearly, the algorithm is beginning to converge on what should be the optimal policy. However, it isn't until after 35,000 iterations that you begin to see any true convergence where the number steps are oscillating through a smaller band.



I hope this helps you guys understand that you don't need to use the MaxDelta parameter to calculate convergence since you can also calculate convergence by the number of steps it takes an algorithm to reach the goal state given a certain number of iterations.

Juan

GAURAV PURI 6 days ago @Juan - The above graphs are for HardGrid Example ? I am getting quite a different graph by running the code

Ben Millice 6 days ago Guarav, I was confused by this at first too, but try changing your y-axis to plot only between 0 and, say, 200, and you'll see the results you were probably expecting.

GAURAV PURI 6 days ago Thank You

Gautam Salhotra 6 days ago I use logarithmic scale for such graph, you can fit quantities that are orders of magnitude higher into 1 graph, and still see the variation.

● Resolved ○ Unresolved

Yingnan Song 7 days ago
Just imported that code into InteliJ IDEA and everything works like a charm. Thanks for sharing this! :)

● Resolved ○ Unresolved

Tyson Bailey 7 days ago
I'm not sure what to tell you guys as I haven't read the code real close, but it's possible to do a set title on some things, so you may want to consider this and then you don't need to have logic to decide when to show things and when not to, you can just pop both of them up. I have linked to the line that I call "setTItle" it doesn't look like you're using the visualizer gui, but I expect there will be something similar.

https://github.com/onaclovtech/burlap-seed/blob/master/src/main/java/AdvancedBehavior.java#L352

Tyson

David Musielewicz 7 days ago So I'm quickly realizing I should just stick to the git and submit pull requests haha.

Just add 'String title' as a parameter to the simpleValueFunctionVis function and then add gui.setTitle(title); just before the gui.initGUI();. and you're set! Now you have your policies titled.

Good find Tyson.

Juan J. San Emeterio 7 days ago Thanks guys. I'll get this merged later today.

Juan

Tyson Bailey 4 days ago git is pretty handy, I was originally hoping the 'burlap-seed' repo would be helpful for others, but glad another one came along that folks found useful.

● Resolved ○ Unresolved

Indira Gutierrez 6 days ago
Are we supposed to modify the GridWorlds when running this experiments?
Is there a place where they explain in detail the grid worlds and why they are interesting? Or are they basically the same as the worlds that are explained in the lecture where you get a score when they go through each of the cells?

Juan J. San Emeterio 6 days ago This is the exact same example described in lecture. It's up to you to figure out why this sort of problem may be interesting in a real world setting.

● Resolved ○ Unresolved

Anonymous 6 days ago
Is there a way to calculate rewards over time ?

Juan J. San Emeterio 6 days ago I just posted this on @762 and I think it answer your question suitably. If you go to that thread they discuss how you can output the reward. Please feel free to modify my code and submit a pull request but consider what I wrote below first since you might not actually have a problem:

"Hey Brent, I too saw Dr. Isbell's post about how the "usual thing" is using the rewards achieved over time. When I originally wrote my code, this question had not been asked and the answer had not been yet given so I used what made sense at the time, which is the number of steps from initial state to the goal, or terminal state.

I just want to make something abundantly clear to anyone reading this that has already plotted out the number of steps from initial state to the goal to measure convergence.
If you did not modify the problem by adding a terminal function with a negative reward: You are fine.
If you only changed the locations of the walls and/or the size of the grid: You are fine.
If you did add a terminal function with a negative reward: You are not fine. You need to modify the output so that the reward function is printed as well since the number of steps might not correlate with the reward function.

For most people using my code with only minor changes, you really should be ok. Just make sure that in your instance of the problem the number of steps and reward function correlate and you should be fine since ultimately the reward function of the MDP I used is simply: Reward Function = (number of steps)*-1 + 100

Thanks,
Juan
PS- If anyone wants to modify the code so that the reward function is printed in addition to the number of steps, please submit a pull request so everyone can benefit as it could be useful for some."

GAURAV PURI 6 days ago Would you discount Future Reward by 0.99 (Gamma Factor) ?

○ Resolved ● Unresolved

Yi Zhou 6 days ago
Thank you so much Juan, this works so well, you rock!

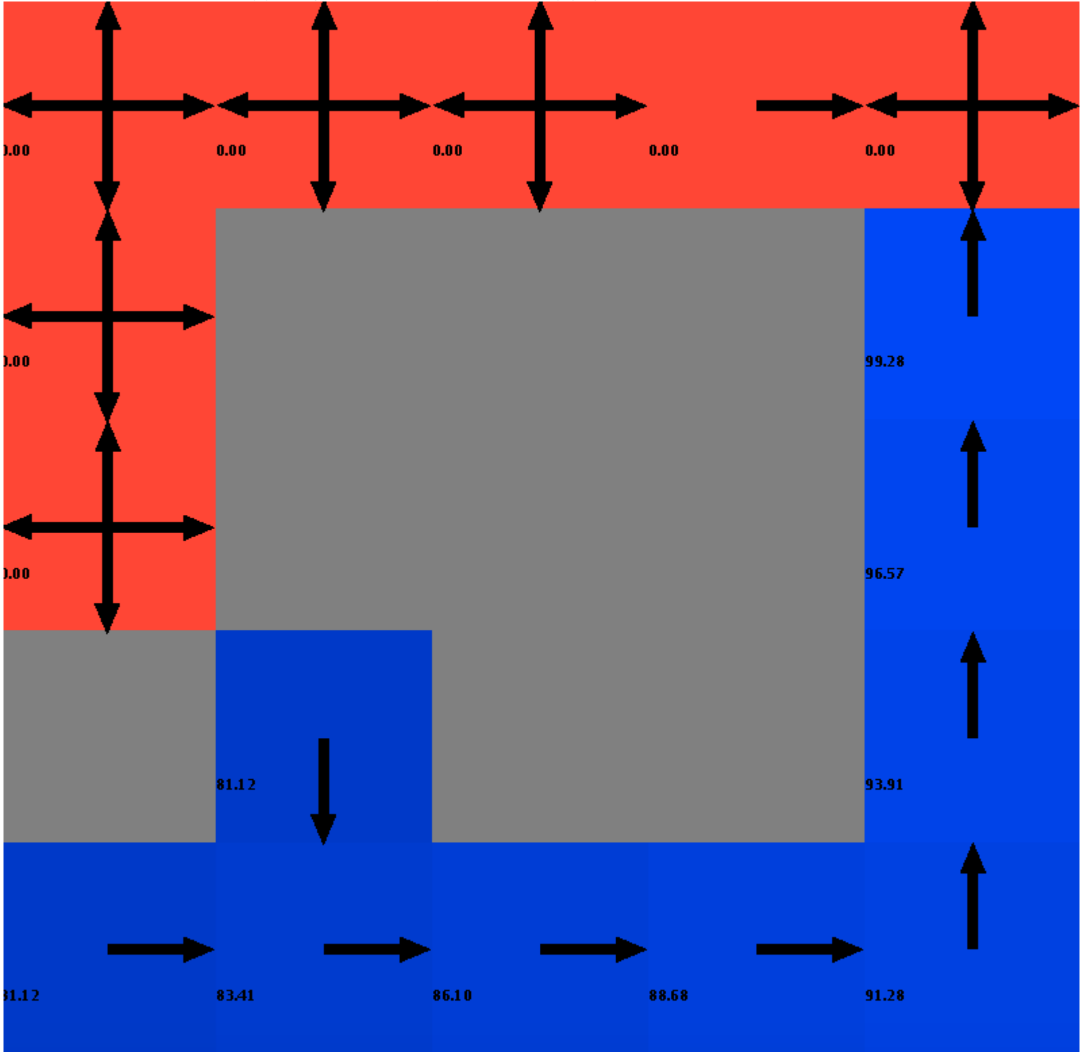
Juan J. San Emeterio 6 days ago No problem, glad you like it!

Yi Zhou 5 days ago Hi Juan,

This might be a stupid question, but I've got this strange answer here:
so when I run your easygirdworld, I got the following result for policy iteration:

```
[>,v,<,>,^]
[^,*,*,*,^]
[<,*,*,*,^]
[*,v,*,*,^]
[>,>,>,>,^]
```

But then when I visualize it using your updated code, I got the following graph:



They don't quite match up with each other, do you know why it behaves like this? Or I miss anything?

BTW, if I want to modify the probability, do I just change the numbers in this part and rerun?

```
public Movement(String actionName, Domain domain, int direction, int[][] map){
super(actionName, domain);
for(int i = 0; i < 4; i++){
if(i == direction){
directionProbs[i] = 0.8;
}
else{
directionProbs[i] = 0.2/3.;
}
}
this.map = map;
}
```

Thank you very much.

Brent Wagenseller 5 days ago Hi Yi,

I am not Juan but I can speculate what is going on.

The top portion of the graph is impossible to reach - Juan's code may just pick a random direction to start, and the colored policy graph shows crosses where it never evaluated (this is why the whole northwest section of the grid is crosses - it cannot reach there from the initial point).

For your second question: I am pretty sure Juan used the initial example from BURLAP as his base. That is how you can change the randomness in the movement (note that is not the epsilon-greedy number, but I guess if you are constrained for time you can modify this for that purpose).

○ Resolved ● Unresolved

Anonymous 6 days ago
Are the passes printed the number of iterations it takes to converge to a solution? for Policy Iteration and Q-Learning.
Also, why does the value iteration don't print the number of passes?

Thanks!

Jeffrey Tagen 6 days ago I admit, I wasn't able to figure out how to determine convergence on Q-Learning. Maybe need to go back to the book.

Maybe just didn't run it long enough on my POS laptop to see it.

● Resolved ○ Unresolved

Salman Faiz Cheem 6 days ago
Juan you rock! Thanks a lot.

Juan J. San Emeterio 5 days ago You're very welcome!

● Resolved ○ Unresolved

Jorge Arguelles 5 days ago
Thanks Juan! This helped a ton!

Juan J. San Emeterio 5 days ago Glad it helped!