

Miniproject 1

Michael Hajostek
School of Electrical Engineering & Computer Science
University of North Dakota
Grand Forks, ND
michael.hajostek@und.edu

I. Q1:

I will be using Obsidian as my primary method of documentation and note taking. I have been using it for a year now as my primary knowledge base so seems rational to keep the majority of my thoughts there. It does leave some to be desired from the product itself, but I haven't found another product that I like better. Other aspects of the projects like how to run it, if ever needed, will be provided in README.md and similar documentation with the source code. I will provide some thoughts in these latex files too.

II. Q2:

I have decided to go with abstract classes and inheritance from that superclass. Think it provides a solid structure as the child class can handle the implementation details but then I can keep the math in the parent class.

III. Q3:

I'm leaning towards element wise rather than matrix form. Matrix form is great for small or dense problems. I think it can be easier to work on in that sense, but the problem is it doesn't scale well. It's $O(N^3)$ notation, which means if anything big is done it collapses completely even with the best hardware.

Element wise on the otherhand is $O(N)$, which is much more manageable and scales nicer. This is a big reason it is used in AI today and the bigger training applications.

IV. Q5:

I think we use not γ^2 because the bellman equation is recursive. So $V(s')$ is only one step away from $V(s)$ so the γ^2 is baked into it already.

Without it becomes an infinite process with positive rewards. The sum would be infinity. We also value immediate rewards more than distant ones.

Usually we use the same but sometimes we have a dynamic discount factor that changes over time.

Some experiments for Q5 with different gammas:

V. Q6:

$$V(s) = \max_a \left[R(s) + \gamma \sum_{s'} P_{ss'}^a V(s') \right]$$

TABLE I
Effect of Discount Factor (γ) on State Values

Gamma (γ)	Goal State Value (V_{goal})	Neighbor State Value ($V_{neighbor}$)
0.1	100.34	3.37
0.5	111.11	22.22
0.9	242.11	157.89
0.99	1741.71	1658.29

To add actions, we can replace the single matrix P with a set of matrix P^a where each matrix corresponds to an action (eg. up, down, left, right). The value convergence method must then change from solving a linear system to solving the Bellman Optimality Equation, as the agent now selects the action a that maximizes expected returns.

VI. Q7:

While the transition logic (P) is factorized and reused to save memory, the Value Function (V) maintains a distinct entry for every global state. Value propagates across room boundaries via the tree structure: when an agent transitions between rooms, the Bellman update reads the value from the destination node (the connected room) and propagates it back to the source node, allowing the agent to 'see' rewards in distant rooms without requiring a monolithic transition matrix.

I would like to play with this more and think it more in terms like retro games like a Gameboy. I watched some videos on them with the techniques they used for the visual palettes swapping them in and out of memory. I think some of those techniques could be useful potentially and a good way to redefine the thinking of the problem itself.

VII. Q8:

I do not like the Markov process model as a whole. Like we talked about in class, it is simple, only relying on the present for any future state. That is great for implementation and helps. But as you talked about too, it only relies on that present state and not the past, which limits the modeling as it doesn't really map to anything in the real world. I can't think of anything whose future doesn't get influenced by the past in some way. Even the one example the one student said about velocity I would push back against that it still relies on the past. I heard

about them in stock prediction too and have seen them have success. They do make sense because the present is all that matters to a degree in stocks but the larger context of the past I think would help with that future state.

Acknowledgment

This material is based upon work supported by the National Science Foundation CISE Graduate Fellowships under Grant No. 2313998. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation.