AI Applications Capstone Project - Maze Solving Agent

Context is all you need

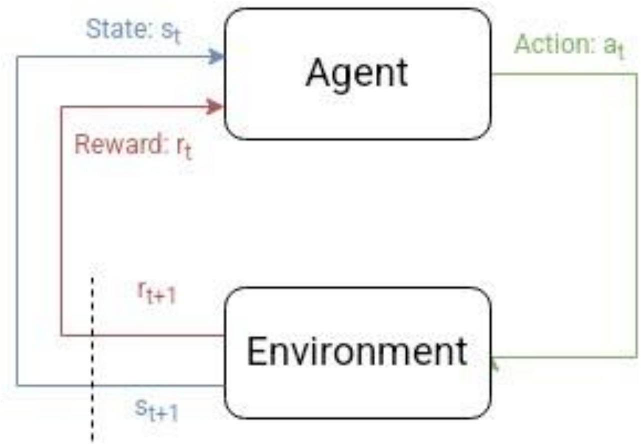
Matthew Thompson - 21st Jan 2025
https://www.linkedin.com/in/thompson-m/

Problem Statement

This project aims to develop and implement a self-navigating agent capable of finding optimal paths through arbitrarily sized maze environments, where complete knowledge of state transitions and rewards is unavailable.

Why read the report and try the code?

- A step into the world of Agents trained by Reinforcement Learning.
- An history of Artificial Intelligence, the input of Animal Psychology, Science of Behaviour, and Reinforcement Learning.
- Lay foundations for further learning about Deep Reinforcement Learning.



Reinforcement learning algorithm

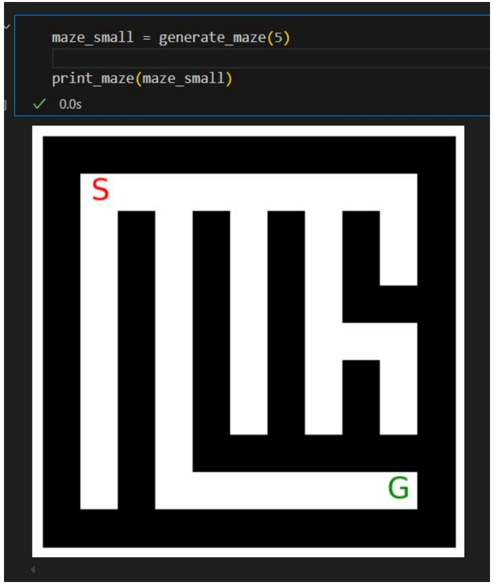
Q-learning (Off-policy Temporal Difference Control) for Estimating $\pi \approx \pi_*$

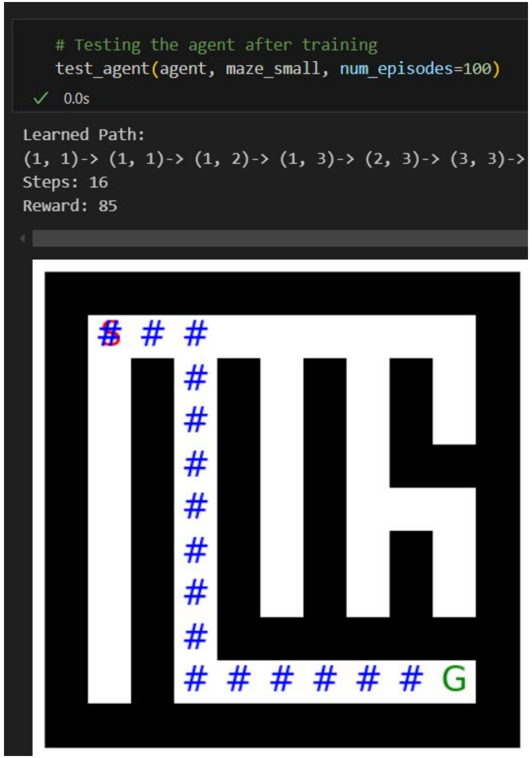
Algorithm parameters: step size $\alpha \in (0, 1]$, small $\varepsilon > 0$
Initialize $Q(s, a)$, for all $s \in S^+$, $a \in A(s)$, arbitrarily except that $Q(\text{terminal}, \cdot) = 0$

Loop for each episode:
1. Initialize S
2. Loop for each step of episode:
   - Choose A from S using policy derived from Q (e.g., ε-greedy)
   - Take action A, observe R, S'
   - Update Q: $Q(S, A) \leftarrow Q(S, A) + \alpha [R + \gamma \max_a Q(S', a) - Q(S, A)]$
   - $S \leftarrow S'$
3. Until S is terminal

What's it do?

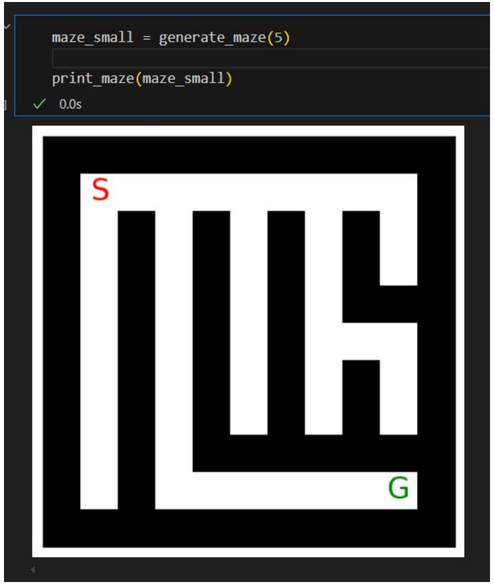
Find the optimal path from the start to the end of the maze !

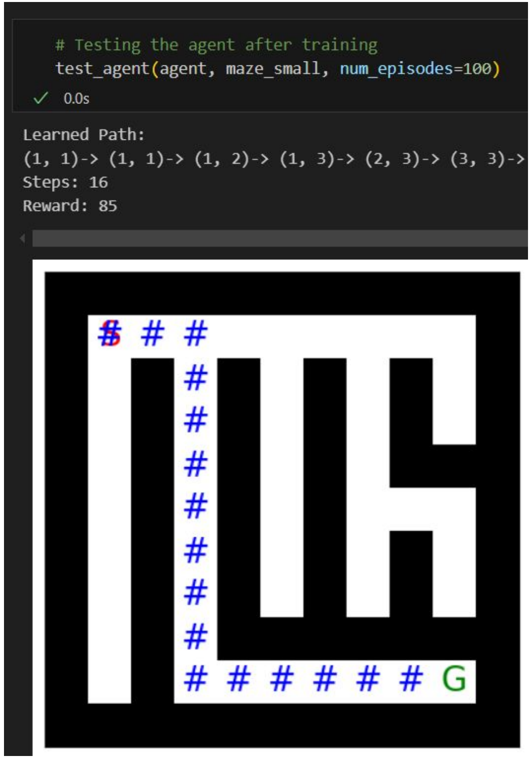




It’s all in the reward

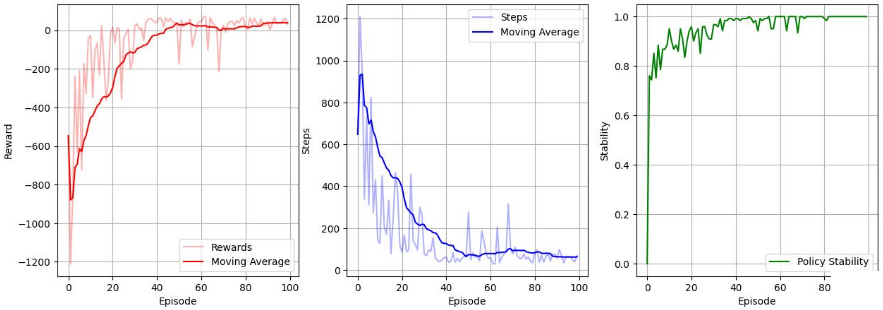
Find the optimal path from the start to the end of the maze !





Training - good times, bad times we all have a share

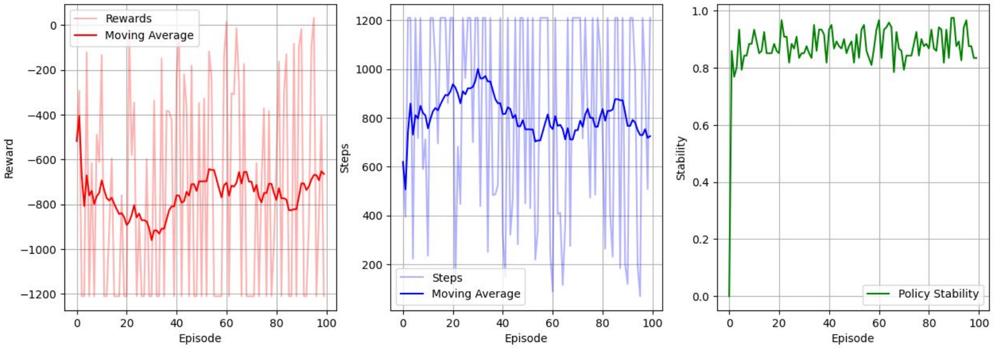
Maze Size: 5x5 | Episodes: 100 | Learning Rate: 0.1 | Discount: 0.99 | Epsilon: 0.3 | Training Time: 103.6s



Good days !

Maze Size: 5x5 | Episodes: 100 | Learning Rate: 0.001 | Discount: 0.99 | Epsilon: 0.1 | Training Time: 106.6s

Bad days !



Key findings and future improvements

The results appear to show the following:

- Learning rate of 0.1 provided optimal for convergence
- Higher exploration rates (epsilon = 3) generally led to faster learning
- The solutions scaled successfully to a 50 x 50 maze
- Training appears to have followed $O(n^2/n)$ complexity

There are other factors that would be interesting to bring in to a future experiment, including:

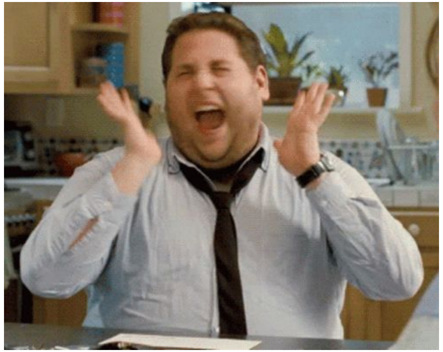
- Experimenting with a higher exploration value
- A metric that tracks when the optimal path is found (i.e. when it converges)
- An early stopping mechanism once it has converged
- Measure the performance of the agents training as the maze scales
- A decaying epsilon value (i.e. high to start with and low as the training progresses)
- Add uncertainty to the environment (e.g. probabilistic wind/slope in some states which move the agent or dirty sensor that gives the agent incorrect location information).

Thank you.

Why this presentation? - Let’s talk about **Capable and Responsible AI Systems**🤓

End of 2022 ChatGPT was launched. Like a lot of people I was blown away. Real live footage of my mornings for the first quarter of 2023

. ------------------------------>>>>>>>>>



Not one but two Mathematical models of Language. Computers now understand and speak our language!!!!

Two you say?

1. Transformer/GPT for sure - models of writing and speech
2. Embeddings - models of the meaning of languages

And the Maths for these are simple.

They understand, however, they also ~~get discombobulated~~ “hallucinate”.

Why do they hallucinate? Is it important to know? Data Protection and security?

Lots of Rabbit Holes here!

I think the Very Smart People* will make inroads over time. It's possible it's a side effect of Neural Networks (humans get things wrong as well… and who hasn't seen the gif of the cat completely failing to make the jump).

My view, LLMs aren't knowledge store, that is a byproduct of understanding language. Though it's a pretty nice by product that the text generated is correct ~80% of the time.

And besides, I work in Security, I have fair to much interest in data protection and integrity to see the LLM as anything but a reasoning tool. My **personal and private data stays separate**.

Which itself has a nice by product……..

Pssst.
Security and Product need the same things from AI Systems/Agents!

*or mice