Workshop2

Brayan Sierra Diaz
20212020036
Daniel Alejandro Presiga
20212020116
Professor Carlos A. Sierra
Systems Engineer
Universidad Distrital Francisco José de Caldas
Bogotá DC

May 9, 2025

1 System Dynamics Analysis

1.1 Mathematical model

In the first workshop, we talked about DQN and how it could be useful for us to improve the capabilities of the agent and obtain better result over time. Now, we propose a math model for the agent behavior similar to a Bellman equation, but for now we think in a equation that approximates the idea of Bellman's equation. The first proposal is this:

L$(\theta) = E_{(s,a,r,s')}\left[(Q(s,a;\theta) - (r + \gamma \max_{a'} Q(s',a';\theta^-)))^2\right]$ Which can be described as:

- $s_t \in \mathcal{S}$ — state of the environment at time step t
- $a_t \in \mathcal{A}$ — action taken by the agent at time step t
- $r_t \in R$ — reward received after taking action a_t in state s_t
- s_{t+1} — next state resulting from action a_t
- $\gamma \in [0,1]$ — discount factor, weighting future rewards
- π — policy that maps states to actions
- $Q(s,a)$ — action-value function: expected return for taking action a in state s and following policy thereafter
- $Q^*(s,a)$ — optimal action-value function
- θ — parameters of the Q-network being trained
- θ^- — parameters of the target Q-network
- $\mathcal{L}(\theta)$ — loss function used to update θ via gradient descent

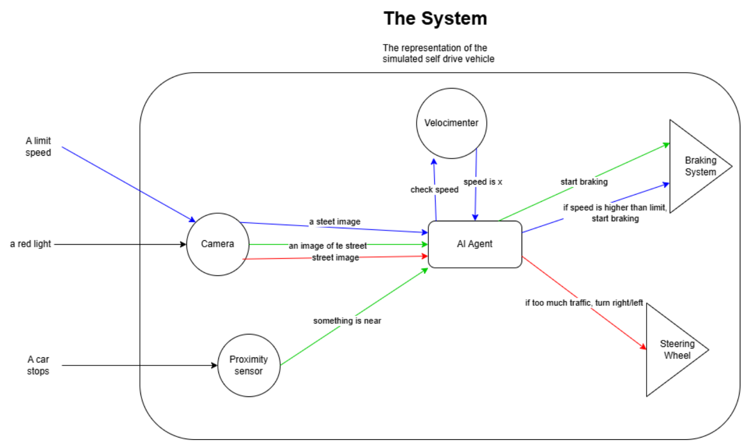
On the other hand, we think that to get a better performance of the agent behavior with DQN, is a good to make the implementation in a discrete simulated portion of a city.

1.2 Phase Portraits or Diagrams

The next image lustrates some of the action and the behavior of the system to the inputs:

On the other hand, the attractors in our case can be:

- A optimal trajectory, because the the agent could converge in one route and not exploring other options, which can lead to a worst behavior over times.
- Another could be a loop route, and therefore not it can not reach the destiny.



For now, we have found that traffic is a chaotic regime, because the action of a car stopping could slow the overall speed average, and as many of traffic actors actions are random, then the traffic becomes chaotic.

2 Feedback Loop Refinement

2.1 Enhanced Control Mechanisms

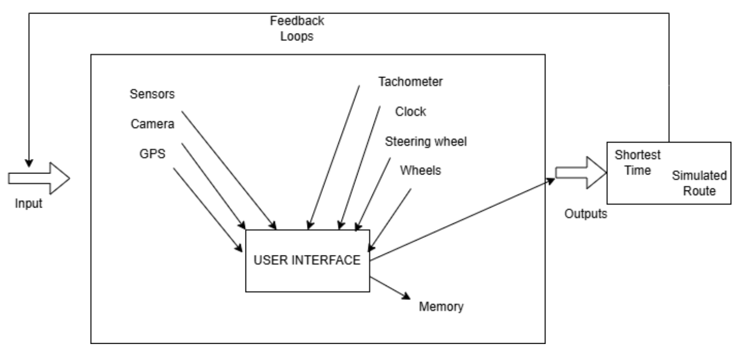


Figure 1: Diagram of the intelligent agent system

As shown in the image, the goal is to implement improved control mechanisms in the system so that it can optimally perceive the environment and react more precisely and efficiently to possible changes:

- A GPS is intended to be implemented, which could provide more accurate and real-time information about the agent's location.
- To improve the granularity of incoming signals, it is proposed that instead of receiving general data such as "obstacles," the system should be able

to process exact distances and other environmental conditions (such as visibility, elements on the road, etc.) that may affect the optimal route and, therefore, the agent's behavior.

For this purpose, the agent is expected to adapt its behavior based on the data obtained, using adaptive control algorithms or reinforcement learning.

2.2 Stability and Convergence

To ensure that the system behaves reliably and improves over time, the following is proposed:

- The agent should exhibit bounded behavior, meaning it must not make illogical decisions such as turning in circles or repeating inefficient routes.
- Through feedback loops, the system should be able to improve future decisions based on past experiences.
- To verify stability, specific metrics could be used, such as:
  - The average time to complete a route.
  - The number of corrections made along the path.
  - Navigation errors detected during the simulation.

3 Iterative Design Outline

- The project plan will be updated to include new data structures and algorithms that optimize the agent's route planning in environments with changing conditions. Additionally, reinforcement learning frameworks will be integrated to enable the agent to learn from the environment and adjust its decision-making strategies, prioritizing the shortest travel time and the selection of the most optimal route.
- The system is intended to be evaluated through simulations, varying some parameters (such as route complexity, weather conditions, etc.) in order to assess efficiency and correct system performance, as well as its ability to adapt to changing conditions, using metrics such as decision-making, problem-solving, and response time.

- https://huggingface.co/learn/deep-rl-course/en/unit2/bellman-equation - https://medium.com/@samina.amin/deep-q-learning-dqn-71c109586bae