Design of A Substantial RL Control System for PULSR2

Intro

The first image below shows how a RL Control System looks generally. The typical RL Control System consists of the agent and the environment. The environment sends two signals to the agent: the observation signal and reward signal. The observation signal tells the agent the current situation of the environment and the agent chooses the most relevant information from the observation signal to determine what the next best action would be; the reward signal helps the agent to know whether the last action taken is a good or bad one. The agent on the other hand influences the environment through its action signal which is as a result of computation based on the received observation and reward signals using the control algorithm that has been deployed in the agent.

In the case of PULSR2, the system (without RL) consists of a control mechanism (control app),DC brushless motors as actuators, robot links as the plant and encoder and load cell as the feedback sensors and data collectors. The image showing how these components come together to form the PULSR2 system is shown among the images below.

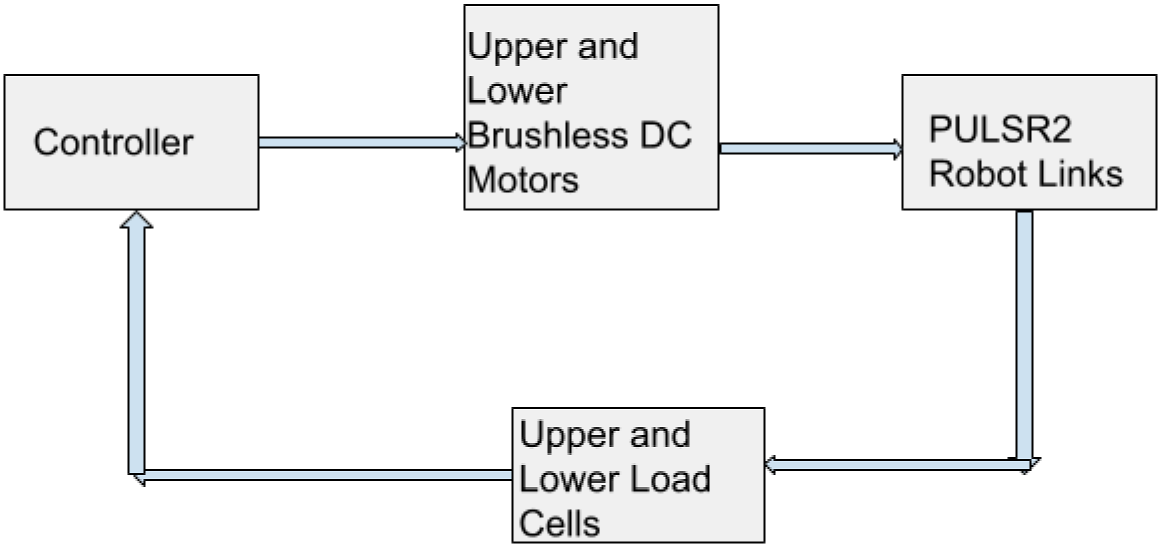
RL Control System for PULSR2

The Agent of the system, where the RL algorithm is deployed and the whole control takes place, is the PULSR2 Control App that is written in Python. This agent produces action signals which are control signals that actuate the DC Brushless Motors and therefore adjust the position and motion of the robot links appropriately. The change in position and motion of the robot links produces some deflections in the load cell and encoder readings which are sent to the agent as the observation signal alongside with the reward of the current (or a previous) action (a scalar value). This process repeats over and over until the end of the session.

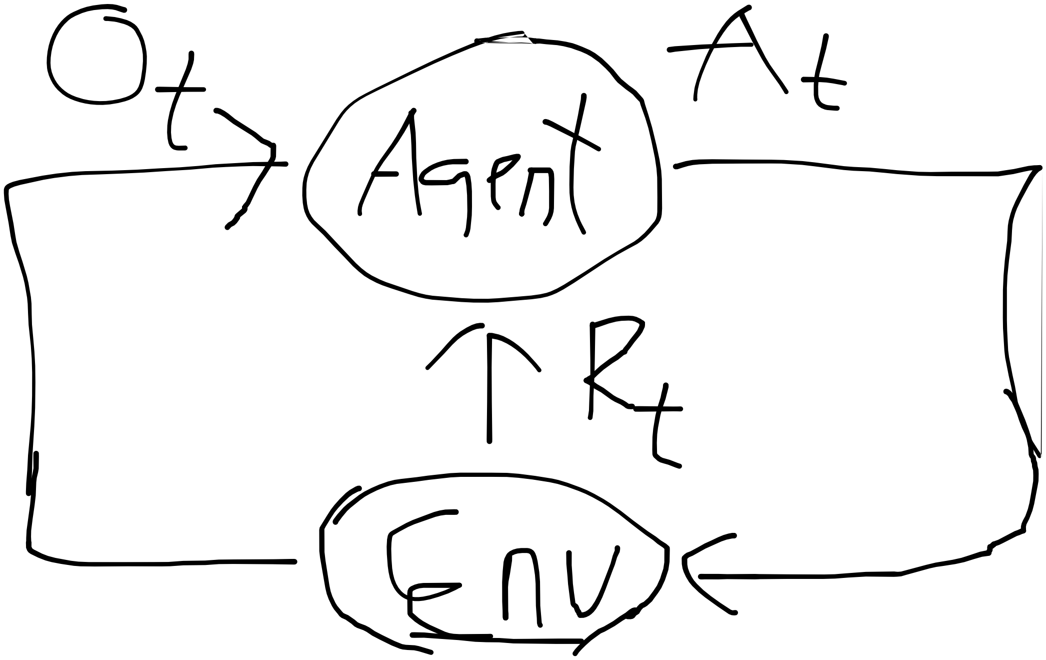
The observation signal is the entire data collected from each action step of the session. The data contains the upper and lower load cell readings, upper and lower motor angle and other important parameters, which sufficiently describe the environment of the PULSR2 during session. The previously acquired data can be used to create an environment model for the agent that it can use to predict the action of the environment. If we do that, we can model the process of determining the best action which is the agent’s major task with a Markov Decision Process (MDP).

The reward signal is a positive or negative number depicting the degree of alignment of the robot with the trajectory. On track maps to positive values while off track maps to negative values.

The substantial design is shown below:


Controller
Upper and Lower Brushless DC Motors
PULSR2 Robot Links
Upper and Lower Load Cells

Components of PULSR2 Robotic System


O_t
Agent
A_t
R_t
Env