Rational Agents

Please do not circulate

The main object of study in this course will be *rational agents*. Rational agents are ones that perform the "best" decisions at any time given. Before we get into the details of the rational agents, it is important to understand the 'concepts' of agent and environment. The conceptual foundation makes it clear to us as to what is that we would want to achieve when we say that we want to design rational agents. A schematic of rational agent is given in Figure 1, where the 4-tuple $\langle S, A, R, T \rangle$ is expanded

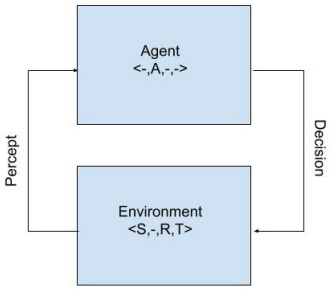

Figure 1: Schematic of Rational Agent

as follows: S for state space, A for action space, R for reward, T to denote the transition. We now further explain what these are. In the textbook, the $\langle S, A, R, T \rangle$ corresponds to **S**ensor (we call it state), **A**ctuators (we call it action), **P**erformance (we call it reward), **E**nvironment (we call it transition T). Out of the 4 tuple, 3 quantities namely S, R, T belong to the environment and only A belongs to the agent (see Figure 1).

State Space The state of the environment is its configuration. Usually, the state is described by *state-variables*. The set of all possible states is the state space S. We use $s \in S$, to denote the fact that the state s belongs to the space S. The state-space could be discrete/continuous, finite/infinite (see Section 2.2) In the case of discrete and finite state space we use the notation $S = \{s^1, s^2, \ldots, s^{|S|}\}$, where $|S|$ is the cardinality (number of elements) of S. The size of the state space grows exponentially with the number of state variables (see Section 2.2).

Action Space The set of all possible actions is the action space A. We use $a \in A$, to denote the fact that the action a belongs to the space A. The action-space could be discrete/continuous, finite/infinite. In the case of discrete and finite action space we use the notation $A = \{a^1, a^2, \ldots, a^{|A|}\}$, where $|A|$ is the cardinality (number of elements) of A.

Reward is usually a real number and is a function of the state and the action denoted by $R(s, a)$.

Transition describes the way the environment changes on applying an action. The transition could be deterministic/stochastic, static/dynamic, episodic/sequential (see Equation (1)).

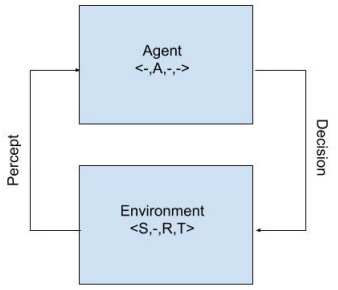

Figure 2: Schematic of Rational Agent

Time	Observation	Action
$t = 1$	o_1	$\pi_1(o_1) = a_1$
$t = 2$	o_1, o_2	$\pi_2(o_1, o_2) = a_2$
$t = 3$	o_1, o_2, o_3	$\pi_3(o_1, o_2, o_3) = a_3$
$\vdots$	$\vdots$	$\vdots$
t	$o_1, \ldots, o_t$	$\pi_t(o_1, \ldots, o_t) = a_t$

Table 1: Shows the use of agent function

1 Hardness of decision making

1.1 What does the agent want?

We saw that the agent only has A at its disposal, and the other 3 quantities namely S, R, T belong to the environment. In Figure 1 this was denoted by the $\langle -, A, -, - \rangle$ (where $-$ is used to stress the fact that the agent does not have those quantities). The percept the agent receives is known as an observation denoted by o_t. The set of all observations is denoted by O, the observation space. The agent needs to map the sequence of observations into sequence of actions. When it is possible for the agent observe the state, i.e., when $o_t = s_t$ we say the agent has *full-observability*, otherwise the agent has only *partial-observability*. For most part of the course, we will only look at the fully observable case.

Agent Function[1] is a map from *states/observations* to the actions. Let us call this π_t. We can make use of π_t in the following way as shown in Table 1.

Let Π_t denote the set of all agent functions at time t, and let π_t^* denote the 'best/correct' agent function. We now update the schematic of the rational agent in Figure 3 by plugging in the observation sequence into the agent function π_t the agent can obtain the action.

1.2 Hardness of computing the agent function

We saw the use of agent function in decision making, we will elaborate on it further. In particular, we will show that $|\Pi_t|$ grows exponentially in the size of the state space (note that the state space grows exponentially in the size of state variables, see Section 2.2). We will consider the case of finite observation space given by $O = \{o^1, o^2, \ldots, o^{|O|}\}$ and finite action space $A = \{a^1, a^2, \ldots, a^{|A|}\}$.

To understand the intricacies of agent function, let us for the moment think about the analogy of a multiple objective quiz. The quiz has $|Q|$ questions and each question has $|A|$ possible answers. Now, the total number of possible answer sheets is $|A|^{|Q|}$ ($=\underbrace{|A| \times |A| \times \ldots |A|}_{|Q| \text{ times}}$). Let us first fix $t = 1$, and measure $|\Pi_1|$. The observation at $t = 1$ is o_1 and it can take one of $|O|$ possible values, i.e., $o_1 = o^1$

[1] The agent function is formally known as *policy*. However, in order to be consistent with the terminology with that of the textbook, we stick to the usage of agent function.

s

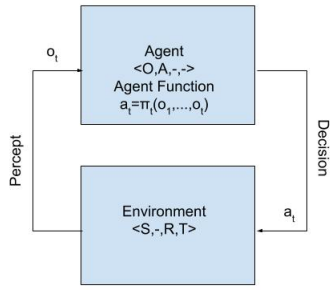

Figure 3: Schematic of Rational Agent

Time	Observation	Action
$t = 1$	o_1	$\pi_1(o_1) = a_1$
$t = 2$	o_2	$\pi_2(o_2) = a_2$
$t = 3$	o_3	$\pi_3(o_3) = a_3$
$\vdots$	$\vdots$	$\vdots$
t	o_t	$\pi_t(o_t) = a_t$

Table 2: Shows the use of agent function for Markovian case

or $o_1 = o^2$ or $\ldots$ or $o_t = o^{|O|}$ (here o_1 the subscript denotes $t = 1$ and $o_t = o^i$ the superscript denotes the i^{th} possible observation from the observation set O). Using the analogy of multiple objective quiz, the agent should be able to act/respond to any one of the $|O|$ possible observations and there are $|A|$ possible actions. Thus, for $t = 1$, we have $|\Pi_1| = |A|^{|O|}$. Now let us extend the case to $t = 2$, now there observation sequence will have two entries o_1, o_2 and the total number of possible observations is $|O|^2$. By similar arguments, we have $|\Pi_2| = |A|^{|O|^2}$.

Thus for general t, we have $|\Pi_t| = |A|^{|O|^t}$. Also, note that the total possible agent functions will be given by $|\Pi_1| \times \ldots |\Pi_t| = |A|^{\sum_{s=1}^{t} |O|^s}$.

One special case arises when the system is Markovian (see Table 2). In this case, the agent does not need the past observations to act in the present. Thus, $|\Pi_t| = |A|^{|O|}$, and the total possible agent functions will be given by $|\Pi_1| \times \ldots |\Pi_t| = |A|^{|O|t}$.

A very useful special case occurs when the environment is *Markovian, Time-Invariant* (see Table 3). Due to Markovian property, given the present the future does not depend on the past, so at time t and a history of present and past observations given by $o_1, \ldots, o_t$, one can discard the past, i.e., $o_1, \ldots, o_{t-1}$ and make the decision only based on the current observation o_t. Thus $|\Pi_t| = |A|^{|O|}$. Further, time-invariance property means that π_t does not change with respect to time, i.e., $\pi_t = \pi$ and hence $\Pi_t = \Pi$, in which case, $|\Pi| = |A|^{|O|}$. **Hardness of decision making** is due to the fact that we have to 'search' for π^* among $|A|^{|O|}$ possible choices of π.

Markovian, Time-Invariant and Fully-Observable

Time	Observation	Action
$t = 1$	o_1	$\pi(o_1) = a_1$
$t = 2$	o_2	$\pi(o_2) = a_2$
$t = 3$	o_3	$\pi(o_3) = a_3$
$\vdots$	$\vdots$	$\vdots$
t	o_t	$\pi(o_t) = a_t$

Table 3: Shows the use of agent function for Markovian and time-invariant case

State	Action
s^1	$\pi(s^1)$
s^2	$\pi(s^2)$
s^3	$\pi(s^3)$
$\vdots$	$\vdots$
$s^{\|S\|}$	$\pi(s^{\|S\|})$

Table 4: Shows look-up table corresponding to the agent function for Markovian and time-invariant and fully observable case

The way to use the look-up table is as follows: at time t, if the state was $s_t = s^i$, then the agent chooses the action given by $\pi(s^i)$, which is in the i^{th} row of the look-up table (see Table 4).

2 Illustrative Problems

2.1 State Variables and State Space

Consider the navigation task in a 2-dimensional room as shown Figure 4. The environment has some blockades. Identify the state variables and state space. Is state space continuous or discrete? Now

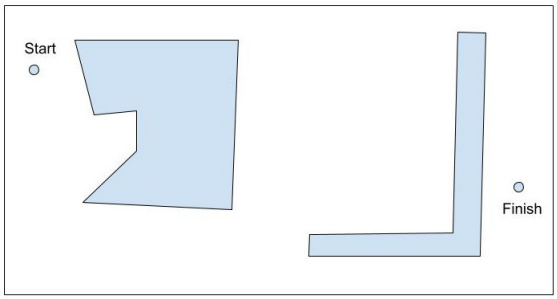

Figure 4: Navigation with blockades

consider different version of the same problem with grids as shown in Figure 5. Let us say there are G grids in each of the dimensions. How does the state space change.

For the above two cases, come up with i) continuous action space and ii) discrete action space.

2.2 Hardness of computing the agent function

Consider a traffic junction environment with 4 roads/lanes. Each lane can have $0, \ldots, L-1$ vehicles. There are 8 lights, 2 for each lane (red and green), and at each instant of time, the traffic signal agent has to decide whether to turn ON the red or the green light. Note that at any given time, only one green light can be ON. For this problem, identify: i) State variables and state space ii) action space iii) size of state space. Also, assume Markovian, time-invariant and full observation setting. Compute $|\Pi|$, i.e., the size of the set of all possible agent functions.

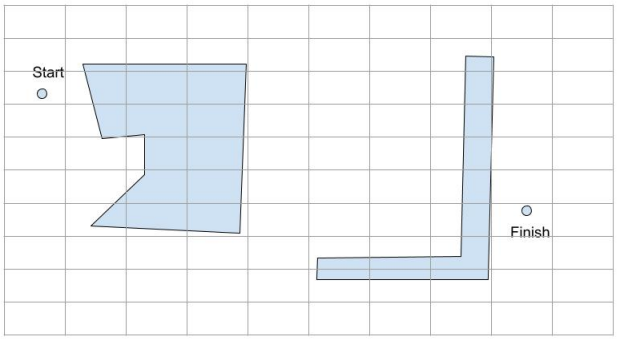

Figure 5: Navigation with blockades

3 Bunny

The bunny is stranded at a random location in the sea and it needs to reach the shore. Assume that the world is 1-dimensional and that the bunny can move one step to the right or one step to the left. Take the state space $S = \mathbb{Z}$ (the set of all integers, where $\mathbb{Z} = -\infty, \ldots, -2, -, 1, 0, 1, 2, \ldots, \infty$). Let the agent be at an initial location s_1 and let the shore begin at a different location s_{shore}

- Write down whether the state space is discrete or continuous, whether it is finite or infinite.
- What is the action space A?
- What is a good reward function? (there can be multiple correct answers)
- What is the transition function T, i.e., identify T such that $s_{t+1} = T(s_t, a_t)$
- What is "best" agent function?
- Give one example of a bad agent function.

Recall the definition of Markovian "given the present, the future does not depend on the past".

4 Particle in a discrete world

A discrete environment contains a moving particle. The dynamics, i.e, transition of the particle is described as follows:

$$\begin{aligned} pos_{t+1} &= \sum_{i=1}^{t} vel_i + pos_1 \\ vel_{t+1} &= \sum_{i=1}^{t} acc_i + vel_1 \end{aligned} \tag{1}$$

where pos_t, vel_t, acc_t are position, velocity and acceleration at time instant t.

- What are the state variables?
- What is the state space?
- Does the transition in (1) depend on history? i.e., does the future depend on present as well as the past?
- Is this system Markovian? If so can you re-write (1) such that the future only depends on the present?

5 Single Stochastic Queue

Assume a queuing environment with no upper limit for the number of customers in the queue. Here, there are two mutually independent events i) Event 1: arrival of a customer who joins the queue with probability $\frac{1}{2}$ ii) Event 2: service of customer who leaves the queue with probability $\frac{1}{2}$. As an analogy, we can imagine this to be like toss of two independent coins (first coin represents the arrival and the second coin represents the service). Note that $prob(HH) = \frac{1}{4}$, $prob(HT) = \frac{1}{4}$, $prob(TH) = \frac{1}{4}$, $prob(TT) = \frac{1}{4}$. Using this analogy:

- What is the state variable?
- What is the state space?
- What is the transition? Say at t we have $s_t = k$ customers, what are the possible next states, and what are the corresponding probabilities?

6 Tasks

1. The environment has a box that contains n red balls and m green balls (say n and m are given and fixed). At time $t = 1$, a ball will be picked at random? The agent has two actions, guess red or guess green. What is the best agent function? How does it depend on n and m?
2. The environment has a box that contains n red balls and m green balls (say n and m are given and fixed). At time $t = 1$, a ball will be picked at random? The agent has two actions, guess red or guess green. The random ball picked at $t = 1$, will now be replaced by a red ball or a green ball with equal probability. However, the agent does not get to see the colors of the replaced ball as well as the replacing ball. At time $t = 2$, the agent again has to make a guess. What is the best agent function? How does it depend on n and m?
3. The environment has a box that contains n red balls and m green balls (say n and m are given and fixed). At time $t = 1$, a ball will be picked at random? The agent has two actions, guess red or guess green. If the agent was successful in guessing the color of the ball, the agent can replace the random ball with a ball of its own choice of color. If the agent was unsuccessful in predicting the color of the ball, the random ball will be replaced by a red ball or a green ball with equal probability, and the agent does not get to see the color of the replaced ball. At time $t = 2$, the agent again has to make a guess. What is the best agent function? How does it depend on n and m?
4. The environment has a box that contains n red balls and m green balls (say n and m are given and fixed). At time $t = 1$, a ball will be picked at random? The agent has two actions, guess red or guess green. After the guess, the agent gets to see the color of the random ball and the agent then replaces the random ball with another ball of its own choice. At time $t = 2$, the agent again has to make a guess. What is the best agent function? How does it depend on n and m?
5. The environment has a box that contains n red balls and m green balls (say n and m are given and fixed). At time $t = 1$, a ball will be picked at random? The agent has two actions, guess red or guess green. After the guess, the random ball is replaced by either a red or a green ball with equal probability. The agent gets to see the color of both the replaced as well as the replacing ball. At time $t = 2$, the agent again has to make a guess. What is the best agent function? How does it depend on n and m?

In all the above cases, identify the categories of the problem:

Problem	Sequential/Episodic	Static/Dynamic	Fully/Partially Observable
1			
2			
3			
4			
5			

Table 5: Mention the appropriate category in the box