Assignment on the Tiger Problem

Ayan Gangopadhyay

May 13, 2020

Introduction

We are trying to find the value vectors that are obtained when we have `LISTEN` as the first action at 2 steps to go.

First we need to determine the number of plans that we can have, this is the *enumeration* part of the process.

Policy for a two horizon for the *tiger problem* is a decision tree which can be represented as shown here.

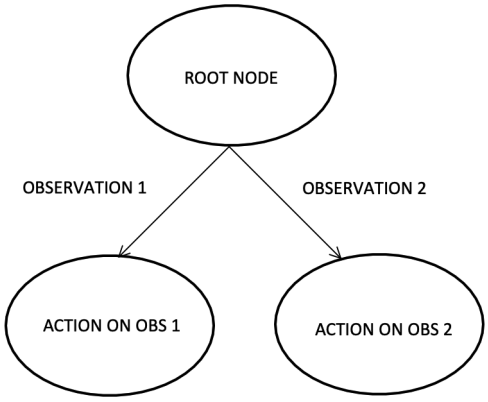

FIGURE 1. GENERAL POLICY TREE

The `ROOT NODE` defines the first action that is taken, as a result of which we see either *observation 1* or *observation 2*, which in the *tiger problem* are *tiger left*(`TL`) and *tiger right*(`TR`) respectively. Further, after we have registered an observation as a result of our action in the horizon with 2 steps *to go*, we have to take another action at 1 step *to go*.

For our problem we are given that the `ROOT NODE is LISTEN.` Since there are 3 actions that can result from either the observation `TL, TR` we have $3^2 = 9$ policies that are possible. Now, out of these policies, only some will dominate at some interval in the belief state (which is binary in the *tiger problem*), the rest are not important and can be discarded. This is the *pruning* part of the algorithm.

Working out a particular policy tree

We now work out the value of the following policy to clearly demonstrate how a policy is evaluated.

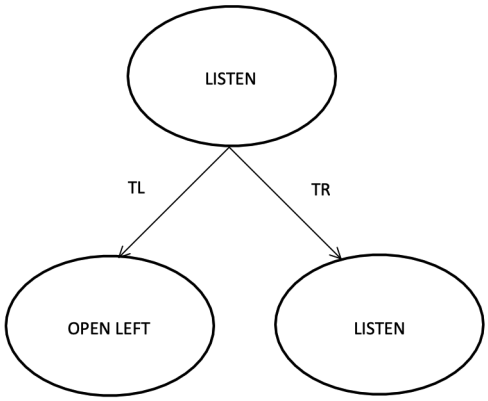

FIGURE 2. EXAMPLE POLICY TREE

The following general formula gives us the value as a function of the current state, where where a_p is the action specified in the top node of policy tree, $o_i(p)$ is the $(t-1)^{th}$ step policy associated with observation o_i at the top level of a t-step policy tree p and we have a time horizon of t.

$$V(s) = R(s, a_p) + \gamma \sum_{s' \in S} T(s, a_p, s') \sum_{o_i \in \Omega} O(s', a_p, o_i) V_{o_i(p)}(s')$$

Rewriting it for our case for the state s, we have:

- 2 steps *to go*
- Our observation space is $\Omega = \{TL, TR\}$
- $T(s, a_p, s') = T$(LEFT, LISTEN, LEFT) *or* T(RIGHT, LISTEN, RIGHT) $= 1$
- $V_{o_i(p)}(s') = R$(s, OPEN LEFT) *or* R(s, OPEN RIGHT) *or* R(s, LISTEN)
- $\gamma = 1$

We need to evaluate the policy for both the states in our problem, these states are *tiger in the left door* $= s_L$ and *tiger in the right door* $= s_R$. Therefore using the above policy tree and values obtained from the *tiger problem* formulation, we have,

$$V(s_L) = -1 + 1 \times (0.85 \times -100 + 0.15 \times -1) = -86.15$$

$$V(s_R) = -1 + 1 \times (0.15 \times 10 + 0.85 \times -1) = -0.35$$

Therefore our value vector will be $\{-86.15, -0.35\}$

Similarly, we will have different different value vectors depending on our policy tree.

Enumeration and Pruning

The value vector for each possible policy are tabulated below and were obtained by running the program attached with this file. This constitutes the *enumeration* part of the algorithm since all possible cases are worked out. This can be a computationally expensive process as is clear from the fact that even for such a simple situation that is described in the *tiger problem* we have to work out 9 vectors for a time horizon of 2.

Action on TL	Action on TR	s_L	s_R
listen	listen	-2.00	-2.00
listen	open_right	-0.35	-86.15
listen	open_left	-16.85	7.35
open_right	listen	7.35	-16.85
open_right	open_right	9.00	-101.00
open_right	open_left	-7.50	-7.50
open_left	listen	-86.15	-0.35
open_left	open_right	-84.50	-84.50
open_left	open_left	-101.00	9.00

Table 1: Value vectors for root node `Listen`

On plotting the value vectors against belief states we get the following plot.

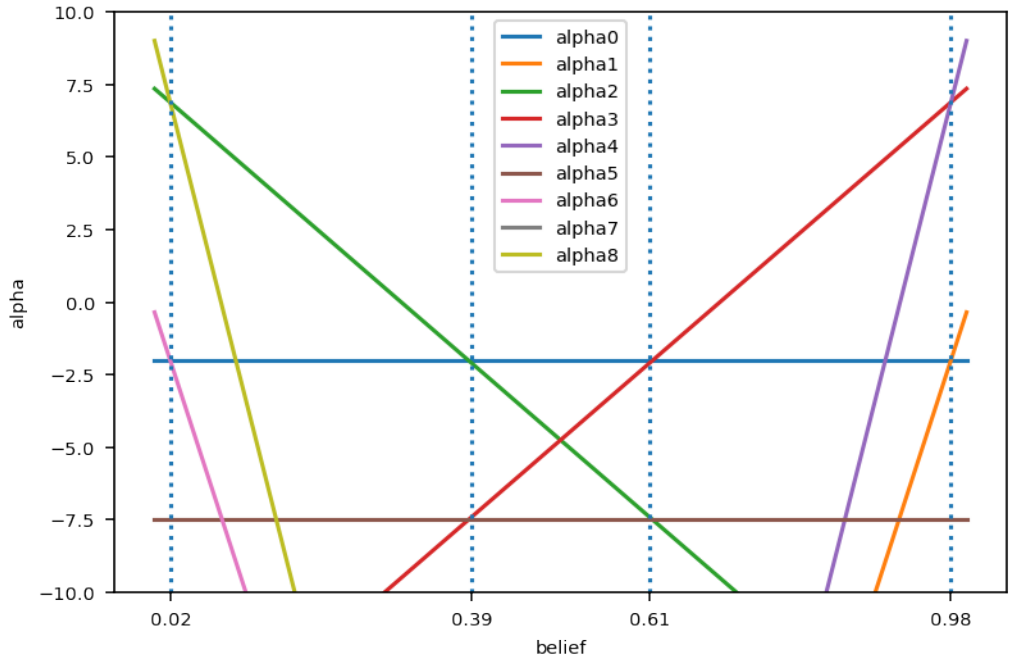

Figure 3. All Value Vectors against Belief

This plot clearly shows that the vectors $\alpha_1, \alpha_5, \alpha_6$ *and* α_7(*which has been excluded from the figure*) do not dominate in any belief state and hence can be dropped or *pruned*.

Finally, the following value vectors were found to be useful for decision making and were plot together.

Action on TL	Action on TR	s_L	s_R
listen	listen	-2.00	-2.00
listen	open_left	-16.85	7.35
open_right	listen	7.35	-16.85
open_right	open_right	9.00	-101.00
open_left	open_left	-101.00	9.00

Table 2: Useful value vectors

Finally we plot the useful value vectors against belief and come up with the following plot.

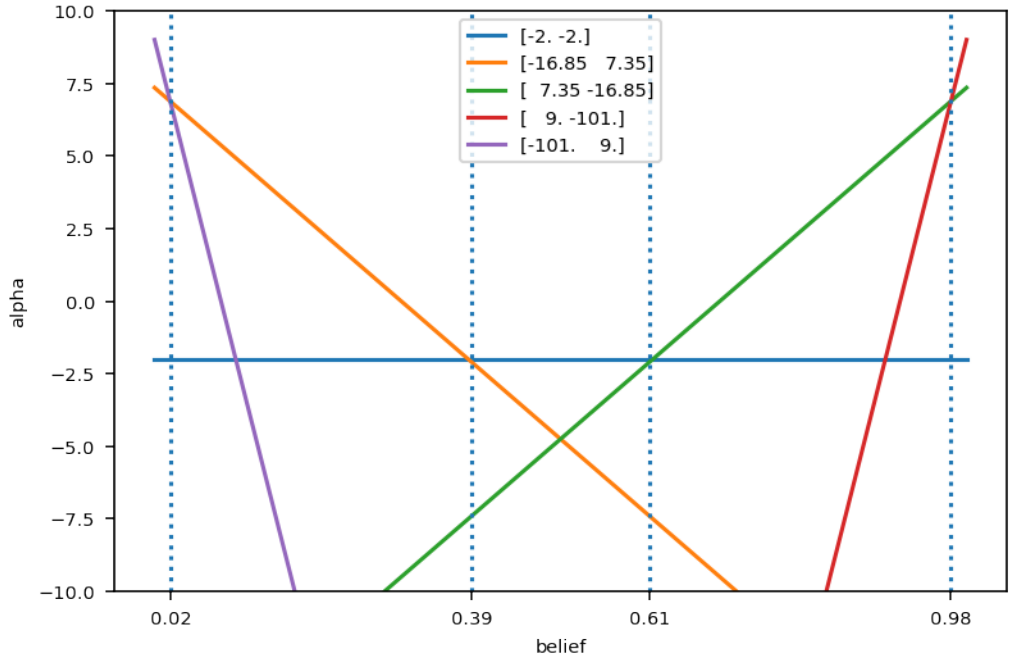

Figure 4. Value Vectors after Pruning