Decision making, Markov decision processes

Solved tasks
Collected by: Jiří Kléma, klema@fel.cvut.cz

Spring 2018

The main goal:

The text presents solved tasks to support labs in the B4B36ZUI course.

1 Simple decisions, Bayesian decision making

Example 1. *(AIMA, 16.10): A used car buyer can decide to carry out various tests with various costs (e.g., kick the tires, take the car to a qualified mechanic) and then, depending on the outcome of the tests, decide which car to buy. We will assume that the buyer is deciding whether to buy car a_1 and that there is time to carry out at most one test t_1 which costs 1,000 Kč and which can help to figure out the quality of the car. A car can be in good shape (s_+) or in bad shape (s_-), and the test might help to indicate what shape the car is in. There are only two outcomes for the test: pass (t_{1+}) or fail (t_{1-}). Car a_1 costs 30,000 Kč, and its market value is 40,000 Kč if it is in good shape; if not, 14000 Kč in repairs will be needed to make it in good shape. The buyers estimate is that a_1 has 70% chance of being in good shape. The test is uncertain: $Pr(t_{1+}(a_1)|a_{1+}) = 0.8$ a $Pr(t_{1+}(a_1)|a_{1-}) = 0.35$.*

Calculate the expected net gain from buying car a_1, given no test.

Use Bayes' theorem to calculate the probability that the car will pass or fail its test and hence the probability that it is in good or bad shape.

Calculate the optimal decisions given either a pass or a fail, and their expected utilities.

Calculate the value of (perfect) information of the test. Should the buyer pay for t_1?

Example 2. *You are going on a trip from San Francisco to Oakland. You have two options to get to Oakland, you want to get there as soon as possible. You can drive your car across the Bay Bridge or go by train through the tunnel under the bay. Bay Bridge is often jammed (on the given part of the day it is in about 40 % of cases). During normal operation, it takes 30 minutes drive. If there is traffic congestion, it takes 1 hour. The train journey always takes 40 minutes.*

When having no traffic information, does it pay off to drive or take a train?

Let us assume, that the traffic information for Bay Bridge is available on web, you can get it in 5 minutes. You know, that for congested bridge, the web page says the same with 90% probability. For normal traffic, the page indicates a traffic jam in 20% cases.

What is the congestion probability when having the traffic information?

What should we do if the traffic information predicts normal operation / congestion?

Is it efficient to spend 5 minutes by finding out the traffic information or is it better to simply set out?

Example 3. *(AIMA, 16.12): Economists often make use of an exponential utility function for money $U(x) = 1 - e^{-x/R}$, where R is a positive constant representing an individual's risk tolerance. Risk tolerance reflects how likely an individual is to accept a lottery with a particular expected monetary value (EMV) versus some certain payoff. As R (which is measured in the same units as x) becomes larger, the individual becomes less risk-averse.*

Assume Mary has an exponential utility function with $R = \$500$. Mary is given the choice between receiving \$500 with certainty (probability 1) or participating in a lottery that has a 60% probability of winning \$5000 and a 40% probability of winning nothing. Assuming Mary acts rationally, which option would she choose? Show how you derived your answer.
Consider the choice between receiving \$100 with certainty (probability 1) or participating in a lottery that has a 50% probability of winning \$500 and a 50% probability of winning nothing. Approximate the value of R (to 3 significant digits) in an exponential utility function that would cause an individual to be indifferent to these two alternatives.

2 Markov decision processes

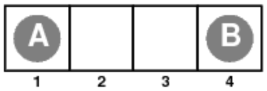
Example 4. *Concern an episodal process with three states* $(1, 2, 3)$. *The rewards in individual states are* $R(1) = -1$ $R(2) = -2$, *and* $R(3) = 0$, *the process terminates by reaching state* 3. *In the states* 1 *and* 2, *actions* a *and* b *can be applied. Action* a *keeps the current state with 20% probability, with 80% probability it leads to transition from* 1 *to* 2 *resp. from* 2 *to* 1. *Action* b *keeps the current state with 90% probability, with 10% probability it leads to state* 3.

Try to guess the best policy qualitatively for states 1 and 2.

Formalize as MDP. Apply policy iteration. Start with the policy $\pi_0 = (b, b, NULL)$ and illustrate its convergence to the optimal policy in detail.

Reapply policy iteration. Start with $\pi_0 = (a, a, NULL)$. What happens? What is the solution?

Example 5. *Consider a two-player game on a four-field board. Each player has one stone, the goal is to move its stone to the opposite side of the board (A player moves from field 1 to field 4, B player from field 4 to field 1). The player that first reaches its goal field wins. The players may move one filed left or right, they cannot skip their move nor move out of the board. If a neighbor field is occupied by the opponent's stone, the stone can be jumped. (example: if A is in the position 3 and B in the position 2 and A moves left, it ends up in position 1).*



Which player wins? Demonstrate the classic solution based on state space search first.

Can we formalize this game as MDP? Is it a good choice?

Formalize this game as MDP. Let $V_A(s)$ be the state s value if A player is on move, $V_B(s)$ be the state s value if B player is on move. Let $R(s)$ be the reward in state s, for the terminal states where wins A it is 1, for the terminal states where wins B it is -1. Draw a state space diagram. Put down Bellman equations for both the players and apply these equations in terms of value iteration. Formulate the iteration termination condition.

Example 6. The tiger problem (POMDP). *An agent stands in front of two closed doors. Behind one of the doors is a tiger and behind the other is freedom. If the agent opens the door with the tiger, the tiger eats the agent (a large penalty -100 is received). If the agent opens the other door, it obtains a reward, its value is +10. Instead of opening one of the two doors, the agent can listen, in order to gain some information about the location of the tiger. Unfortunately, listening is not free, it costs -1; in addition, it is also not entirely accurate. There is a 15% chance that the agent will hear tiger behind the left-hand door when the tiger is really behind the right-hand door, and vice versa. If the agent listens to the tiger repeatedly, it has to pay its cost repeatedly, however, the mishearings can be considered independent.*

Formalize the tiger problem as a partially observable Markov decision process.

Find the optimal 1-step plan as a function of belief. In other words, propose the optimal action as a result of your belief in the tiger's current location. Identify the belief space positions where the action changes.

How many conditional plans of length 2 do we have? Calculate the utility of one of them (it is a function of b again). Will be any of these plans clearly dominated by the other plans?

How many times does the agent have to hear the tiger from the same direction to open a door? Consider the beginning of the game where $b(TR) = 0.5$. Explain.