Homework 3 – AI Algorithms

Noémie MAZEPA

Lorrain MORLET

Table of Contents

Introduction on MDPs:

Before focusing on any example, let’s start by giving a quick overview of what a Markov Decision Process is and how it can be applied.

Markov Decision Processes were invented at the beginning of the 1950s and take a great inspiration from Andrey Markov’s work on stochastic processes. They are a mathematical framework used for decision-making in an environment where the outcomes are influenced by the actions of an agent as well as probabilities.

MDPs are composed of:

- a **set of states** which represent all the possible configurations of the environment.
- a **set of actions** which are the choices available to the agent in each state.
- a **transition function** which specifies the probability of transitioning from one state to another given an action.
- a **reward function** which gives a numerical value as feedback after a transition to tell the agent if reaching a particular state or taking a specific action was a good thing or not.
- a **policy** which is a strategy that defines which action is best to take in each state. To get the optimal policy, we choose the one that maximizes the total reward iteration after iteration.
- a **probability distribution** which describes the likelihood of starting in a certain state or be in certain states during the process.

The main goal of a MDP is to maximize the total of rewards over time. The more an agent gets rewards, the more likely it is to reach the wanted goal. MDPs are especially useful for problems that require sequential decisions over because they provide a good balance between short-term rewards and long-term objectives. For example, an agent might take actions that seem less optimal at first but end up with better rewards later.

Today, Markov Decision Processes are used in several domains like robotics, video games, economics, and artificial intelligence. They can solve a lot of different problems thanks to their flexibility such as controlling self-driving cars or optimizing healthcare treatments. This adaptability makes them a very important tool for reinforcement learning and predictive modeling, where agents make decisions to meet specific goals or need to learn optimal behaviors with feedback from their environment.

We also wanted to mention that MDPs play a big role in artificial intelligence, especially in the generation of images, text or music although those tasks sometimes use similar ideas like Markov Chains. Indeed, in image generation, MDPs help split the image-creation process into smaller steps. Each step involves deciding how to change things like pixels, colors, or shapes. The reward system guides the AI to create better or more specific images by constantly checking and improving each step. This demonstrates just how flexible MDPs can be, even for creative and dynamic tasks.

1st example: Restaurant seating

For the first example, we chose to set the context in a busy restaurant where the manager needs to figure out the best way to seat customers to maximize their satisfaction and his revenue. Managing a very busy restaurant during peak hours is complex and it requires making good decisions to ensure great customer service and profitability. The biggest challenge is seating groups of customers efficiently while managing limited tables, different group sizes, and fluctuating wait times.

This situation is perfect for using a Markov Decision Process (MDP), which helps make step-by-step decisions in an unpredictable environment with many factors to take into consideration.

In this example, we will imagine a restaurant with several table configuration like tables for two, four or six people. During the busy periods, groups of customers arrive randomly, and they all have different sizes, patience levels and spending habits. The manager needs to decide which groups to seat, when to combine or split tables, and whether to ask a group to wait or even turn them away. Each decision affects not only immediate gains, like revenue from seated customers, but also long-term goals, like keeping customers satisfied and building loyalty.

Let's define the states, actions, rewards and transition function that will compose our Markov Decision Process:

States:

They represent the current situation of the restaurant. We have:

- **Table availability**: the information on which tables are free and how many seats are available.
- **Customer groups**: The size of groups waiting for a table and how patient they are.
- **Current wait time**: How long these groups have been waiting for.
- **Restaurant occupation**: The total number of customers currently seated.
- **Revenue so far**: The total money earned during the shift.
- **Customer satisfaction levels**: How happy the seated and waiting customers are.

Actions:

The manager has several actions to choose from in each state:

- **Seat a group:** Assign a group to an available table (this might involve combining or splitting tables).
- **Ask a group to wait:** Decide not to seat a group right away.
- **Turn a group away:** Refuse service to a group if there are no suitable seating options.
- **Combine tables:** Rearrange smaller tables to create larger ones for bigger groups.

- **Split tables:** Divide larger tables to fit smaller groups.
- **Prioritize a group:** Move a specific group up in the line based on their size, patience, or spending habits.

Rewards:

The rewards measure the results of each action, considering immediate and long-term benefits as well.

- **Revenue from the seated customers**: The money spent by groups that are successfully seated.
- **Customer satisfaction**: Positive feedback from reducing wait times or seating customers quickly.
- **Penalty for dissatisfaction**: A negative reward if groups leave because of long wait times or if they were turned away.
- **Fidelity impact**: A bonus reward for keeping customers happy and encouraging them to return.
- **Efficient use of tables**: A reward for using tables effectively and minimizing wasted space.

By using an MDP approach, the restaurant can create a seating plan that adjusts to changes quickly. This way, the business can make sure everything runs smoothly, even during busy times. This example also shows how MDPs can handle tasks with specific rules, where multiple goals need to be balanced in a changing and unpredictable setting. It also shows how important it is to balance short-term and long-term rewards to achieve the best results: maximizing revenue and customer satisfaction.

2nd example: Soccer penalty kicks

For the second example, we're going to talk about penalty kicks in soccer, where the goalkeeper needs to decide how to save a shot under huge pressure. Penalty kicks are crucial especially during close matches, where saving the goal could completely change its outcome. The goalkeeper has a serious challenge: choosing the right action (going left, going right, or staying in the center) without knowing exactly what the penalty taker will do. We can also add that the goalkeeper's actions and previous decisions can influence the kicker's future behavior. This scenario requires a mix of strategy, intuition, and adaptability and it's why it is a great example for the Markov Decision Process.

Although intuition and chance play a certain part in the goalkeeper's final decision, he can also factor in things like the kicker's habits, body language, and even the overall stakes of the match (if the adverse team is winning or losing for example). In this situation, MDPs help break it down into states, actions, and rewards, therefore making it easier to figure out the best strategy. By using an MDP, we can explore how goalkeepers balance risk and reward, adapt to the other's behavior, and make decisions that give them the best chance to save the goal.

Let's see what the states, actions, rewards, and transitions are in this scenario:

States:

The states represent the current context of the penalty kick situation.

- **Position of the kicker**: His body language and the angle he chose to direct his kick.
- **Previous kicker behavior**: Past information about the kicker's habits like his usual direction to shoot.
- **Stakes of the game**: The current score and importance of the penalty (if we are in the regular time, extra time, or penalty shootout).
- **Goalkeeper's past actions**: The directions the goalkeeper chose in previous penalties.
- **Kicker's confidence level**: An estimate based on the match context or hesitation in the movements.

Actions:

The goalkeeper can choose to do several actions during a penalty kick.

- **Dive left**: Move to the left side of the goal.
- **Dive right**: Move to the right side of the goal.
- **Stay center**: Stay in the middle without moving.
- **False move**: Pretend to move early to influence the kicker's decision.

Rewards:

The rewards show the results of the goalkeeper's actions.

- **Save the penalty**: A positive reward for successfully stopping the ball.
- **Concede a goal**: A negative reward for failing to save the shot.
- **Psychological advantage**: A bonus reward for making the kicker hesitate or mess up during future penalties.

- **Match impact**: A bigger reward or penalty based on the importance of the kick. For example, if it was a deciding penalty or a regular shot.

- **Consistency bonus**: A reward for correctly guessing the kicker's habits multiple times.

By using the MDP approach, the goalkeeper can create a plan that quickly adapts to the kicker's actions and the game situation. This helps them make important saves, even in high-pressure moments. This example shows how MDPs can be used for tasks that need quick decisions, deal with uncertainty, and balance different goals. It shows the importance of both short-term actions, like choosing the right direction to dive, and long-term plans, like influencing the kicker's future behavior. This helps achieve the best results and gives the team the best chance to win.

3rd example: Music composition

For our third example, we wanted to look at a more creative and abstract scenario: composing a piece of music. We imagine an AI composer that needs to create a melody that not only sounds harmonious but also keeps listeners interested and emotionally connected. Each note or rhythm chosen must fit well with the previous ones to create a coherent piece of music. This step-by-step decision-making process, where each choice affects the final outcome, makes music composition a great example of using a Markov Decision Process (MDP).

The challenge is to find the right balance between structure and creativity. Every decision like repeating a theme, introducing a new one, or changing the rhythm has its pros and cons. The "reward" here could be listener satisfaction, measured by how emotionally resonant, engaging, or novel the music is.

What makes this example particularly interesting is that it pushes the boundaries of MDPs. In this scenario, it's about dealing with abstract concepts like aesthetics and emotional impact. By using MDPs, we can model this creative process as a series of decisions, where the AI learns to optimize for both technical accuracy and artistic quality.

Let's look at the states, actions, rewards, and transitions that define this MDP:

States:

The states show the current situation of the music being created.

- **Current musical structure**: The melody, the harmony or the rhythm at the moment. For example, the key, the speed and the style.

- **Emotional tone**: The mood of the music so far such as happy, sad or aggressive, etc.

- **Repetition**: If parts of the music have been repeated or changed.

- **Engagement levels**: Feedback from listeners or models that show how people might react.

Actions:

The actions are what the AI can choose to do at each step.

- **Add a new note or chord**: Pick the next music note or chord based on the current state.
- **Repeat a theme**: Use a part of the music that was already played.
- **Change the rhythm**: Change the speed or timing of the music.
- **Change the emotional tone**: Move to a different mood. For example, from happy to sad or aggressive to calm.
- **End the piece**: Finish the music when it "feels" done.

Rewards:

The rewards show how good the music is at each step.

- **Listener satisfaction**: A positive reward based on how well the music connects with the listener's emotions.
- **Uniqueness**: A reward for adding unique and interesting parts to the music.
- **Coherence**: A reward for keeping the music consistent throughout the whole piece.
- **Penalty for dissonance**: A negative reward for parts that sound bad or don't fit well.
- **Finalization reward**: A final bonus for making a piece that feels complete/finished.

By using an MDP approach, the AI can make sequential choices that need creativity and emotional impact. This helps create a piece of music that is melodic and pleasing, and that connects with listeners. This example also shows how MDPs can handle a tricky and creative job like composing music, where each choice adds to the previous ones to create the final piece. It needs to balance between short-term goals like picking a nice-sounding note, and long-term goals like making a song that's emotionally powerful.

By using MDPs for music composition, we see how the model can deal with abstract and artistic challenges, showing its flexibility beyond just structured problems. This adaptability makes MDPs a powerful tool for both technical and creative tasks. It is great for step-by-step decisions even in areas where personal taste and aesthetics are the most important elements. It also shows how AI can boost human creativity by offering new perspectives on artistic choices.

Conclusion

In this homework, we explored the versatility and power of Markov Decision Processes (MDPs) through three diverse examples: restaurant seating, soccer penalty kicks, and music composition. Each scenario highlighted the adaptability of MDPs in handling complex decision-making processes across different domains.

1. **Restaurant Seating**: In a bustling restaurant, MDPs helped figure out the best way to seat customers to keep them happy and boost revenue. By breaking down the situation into states, actions, and rewards, the manager could make smart choices that balanced immediate needs with long-term goals, like keeping customers loyal and using tables efficiently.

2. **Soccer Penalty Kicks**: During high-stakes penalty kicks, MDPs gave goalkeepers a structured way to decide their moves based on the kicker's behavior and the game situation. This showed how MDPs can deal with uncertainty and quick decisions, helping goalkeepers make the best choices under pressure.

3. **Music Composition**: In the creative process of composing music, MDPs helped model the step-by-step decisions needed to create a beautiful and emotionally engaging piece. This application highlighted how MDPs can handle abstract ideas like aesthetics and emotional impact, making them a powerful tool for both technical and artistic tasks.

Through these examples, we saw that MDPs can be used in a wide range of problems, from structured tasks to more creative and dynamic ones. Their ability to balance short-term rewards with long-term goals makes them super useful in fields like robotics, economics, and AI.

In a summary, Markov Decision Processes are a flexible and powerful tool for making decisions in uncertain environments. They help us achieve specific goals and optimize behaviors in complex situations, making them a go-to method for reinforcement learning and predictive modeling.

Sources

https://www.spiceworks.com/tech/artificial-intelligence/articles/what-is-markov-decision-process/

https://en.wikipedia.org/wiki/Markov_decision_process

https://www.google.com/search?client=firefox-b-d&q=the+markov+decision+process#fpstate=ive&vld=cid:5898cfbe,vid:e09zIakMwGU,st:0

https://www.google.com/search?client=firefox-b-d&q=the+markov+decision+process#fpstate=ive&vld=cid:5898cfbe,vid:e09zIakMwGU,st:0

https://medium.com/after-the-full-time-whistle/using-markov-models-to-analyse-defensive-strategies-in-modern-football-simplified-explained-68097ac008e8

https://www2.it.uu.se/edu/course/homepage/aism/st11/MDPApplications3.pdf

https://saturncloud.io/glossary/markov-chains-in-generative-ai/

https://towardsdatascience.com/markov-chain-for-music-generation-932ea8a88305

https://etheses.whiterose.ac.uk/13286/1/Thesis%20Final%20Corrections%20MG.pdf