Markov Chain Example

Reinforcement Learning Study

2021-02-01

차 례

강의현(Lazy Minsu)

Lazy Minsu spends most of his day in bed. Minus's mom nags at the lazy Minsu. "Don't just lie down like that. Sit down and study!". Minsu thinks about getting out of the bed after hearing mother's nagging. Minsu has no choice but to sit at the desk. But, Minsu has no patience. The probability that Minsu still sits at the desk after one hour is only 0.2, and once he goes back to bed, the probability that he sits at the desk again is 0.5. What will Minsu become when he grows up?

1. What is the state space?
2. What is the transition matrix?
3. What is the initial distribution?
4. After 8 hours, where is Minsu more likely to be?

권도윤(University admission)

In the Dark Ages, Harvard, Dartmouth, and Yale admitted only male students. Assume that, at that time, 80 percent of the sons of Harvard men went to Harvard and the rest went to Yale, 40 percent of the sons of Yale men went to Yale, and the rest split evenly between Harvard and Dartmouth; and of the sons of Dartmouth men, 70 percent went to Dartmouth, 20 percent to Harvard, and 10 percent to Yale. We form a Markov chain with transition matrix

1. What is its state space?

$$S = \{H, D, Y\}$$

2. What is its transition matrix P?

$$P = \begin{pmatrix} 0.8 & 0.2 & 0 \\ 0.3 & 0.4 & 0.3 \\ 0.2 & 0.1 & 0.7 \end{pmatrix} \tag{1}$$

3. Is the Markov chain periodic or aperiodic? Explain and if it is periodic, also give the period.

aperiodic

4. Find the stationary distribution.

$$(\frac{5}{9}, \frac{2}{9}, \frac{2}{9})$$

권태현(Son-heung-min score)

Son Heung-min is a South Korean football player and one of the best strikers in the EPL. In particular, Son Heung-min showed strong performance against the opposing team wearing a yellow uniform in the 2020 season. On average, he scored 1.5 goals per game against teams in yellow uniforms. On the other hand, he scored 0.7 goals per game for opponents in different colors. The team in the EPL is 19 teams except Tottenham, where Son Heung-min belongs, and each team will play 38 games. Son Heung-min is famous for scoring many goals in a short period of time in the league. When he scored in the previous game, he had a 0.8 chance of scoring in this match. If you don't score a goal in the previous game, you have a 0.6 chance of scoring in this game. Son Heung-min has scored three goals in seven games, and the team he will meet this time is Wolverhampton wearing a yellow uniform. When Son Heung-min finishes this season, what is the expected number of points?

*state : the team of match with tottenham

*transition probaility : 0.8 0.2 / 0.6 0.4

*reward : goal

Solution

1. Find the transition matrix
2. Show the markov chain
3. Find the expected reward

김봉석

Introduction (Markov Chain Applicaition in Personalized Recommendation System)

Recommendation system is one of the hot topics these days. Companies like Netflix and Watcha are making great success using the recommendation system as a key technology. as I know, they are using item-based or user-based recommendation systems or Hybrid Ensemble Technique. But Usr's web log data or Purchase data contians TimeSeries history, which is one of most important feature in prediction of future works. I Think it would be modeled as a Markov Chain, further more MDP or RL Problem

We will be able to answer 'what is the best recommendation item or what is optimal action in now ?' in the process of solving this problem.

Unlike User-based and Item=based Recomendation System, I want to create a more personalized agent.

Problem Formulation

There is a clear limit because I do not know the Makov chain perfectly yet. Following, the basic level of Problem Folmulation I tried is described.

I'll give you an example MDP with Netflix that I like.

- State : set of recent history of Watching Movies. ex) recent 5 history of recent Movie Title and Raiting
- Action : Based on that history(State) recommend several similar movies, which will be action of each of Recommendation
- Trainsition Probability : ? 1 or 0
- Reward : if User see the Recommended Movies +1 else 0 or Rating Point

we will find what is Optimal action to maximize the reward,

the agent will be traind, The more rewards receive, the better will recommend

FInally, The more users use, the more likely have a better personalized agent.

Imitation

yet, I only came up with ideas, but I do not know how to model them perfectly.

what is Transition Probability? .. State?

박재민 (Personal device prediction)

There is 2 big personal device maker. Apple and Samsung. Both company has their own phone and tablet products. Customers can use two personal devices as the same company's products or they can mix them. Since using same company devices is more compatible between devices, probability of using same company device after 1 year is 0.9. It is uncomfortable to use phone and tablet as different company, there are more probability to make devices as same company. Detailed probabilities are shown in the diagram.

Apple's devices are slightly more expensive than Samsung, so using both Samsung products cost 100, both Apple 120, mixed 110 per year.

Assume that you're early adopter, you must change your phone and tablet every year.

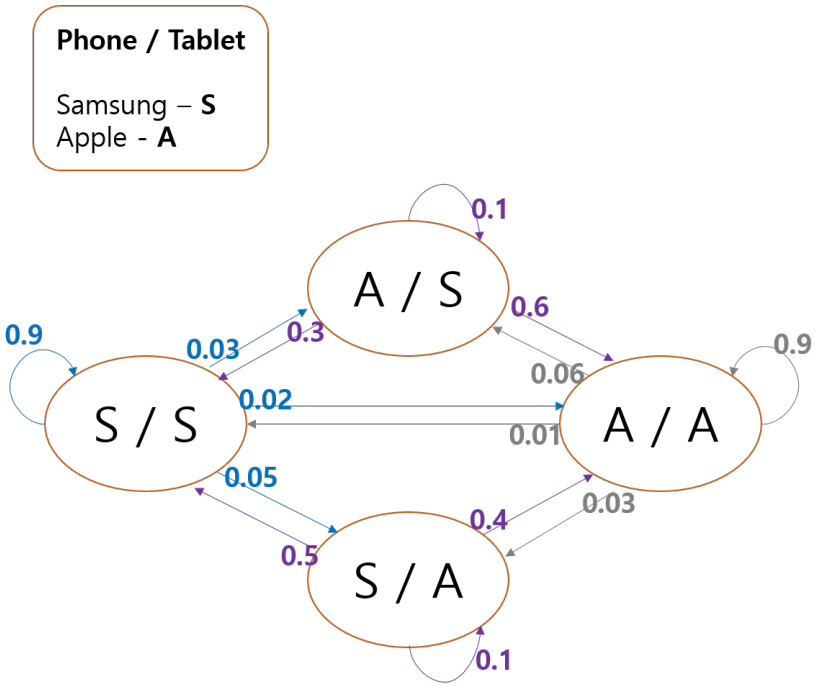


그림 1: transition

State space - {S/S, A/S, S/A, A/A}

Reward - {-100,-110,-110,-120}

Transition Matrix

$$P = \begin{pmatrix} 0.9 & 0.03 & 0.05 & 0.02 \\ 0.3 & 0.1 & 0 & 0.6 \\ 0.5 & 0 & 0.1 & 0.4 \\ 0.01 & 0.06 & 0.03 & 0.9 \end{pmatrix} \tag{2}$$

Questions

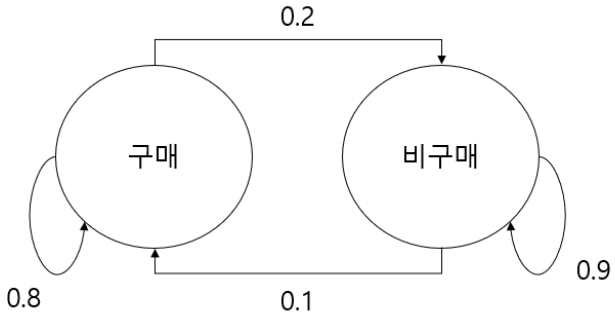
1. If I'm using Samsung phone and tablet, after 10 years, what is the probability of using both Apple devices?

2. If I'm using Apple phone and tablet, How much on average I spend on personal device for 20 years?

3. What is stationary distribution?

백종민 (Predict Consumer)

There is an Internet shopping mall site. Among the members, 30,000(3) consumers purchased in January and 10,000(1) non-purchasing consumers. Consumers who purchased products in January are 0.8. Those who did not purchase products in February are 0.1, and those who did not purchase products in January are 0.1, and those who did not buy products in January are 0.9.



1. what is the state space, transition matrix p, initial distribution.

2. Is the Markov chain periodic or aperiodic?

3. Forecasts of consumers' purchases and non-purchases in March.

4. Forecasts of consumers' purchases and non-purchases in April.

5. What is the probability of consumers purchasing and non-purchasing after n months(limiting probabilities)

6. Use eigenvector and eigenvalue to get a fixed distribution.

손민상(Effect of advertisement for new product)

Introduction

When the company decide whether they advertise some product or not, they compare the effect of the advertisement. And if they do not have a promotion for the product, the product will lose its popularity. The effect of the advertisement is also decided by the type of the advertisement.

Problem

MS company release new product and plan to promote for this product. There are three choices: Youtube advertisement, SNS advertisement, do nothing. State 1 means the people who do not know the product or be willing to buy the product; 2 means the people who sometimes buy the product; 3 means regular customers. The diagram for each effect is below. Make an optimal policy.

- $State = S = \{1, 2, 3\}$
- $Reward \quad R(1) = 0\,, R(2) = 100\,, R(3) = 1000$

If advertisement cost and number of customers are given, we can make a optimal problem. and we also have to think about advertising using both Youtube and SNS in same time. I think it need to think about differcy of results from different duration.

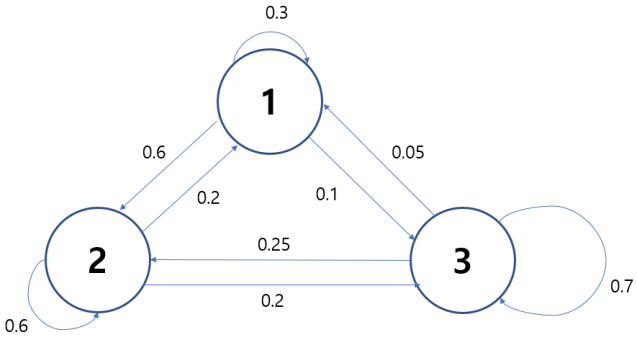


그림 2: Youtube

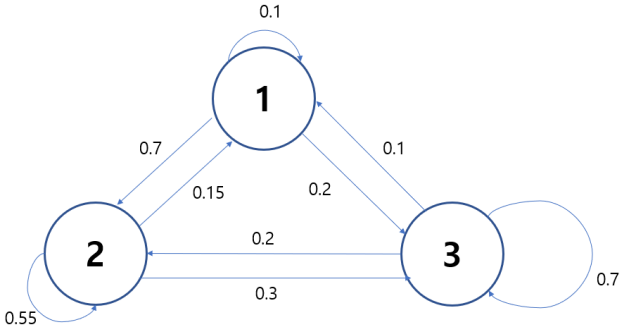


그림 3: SNS

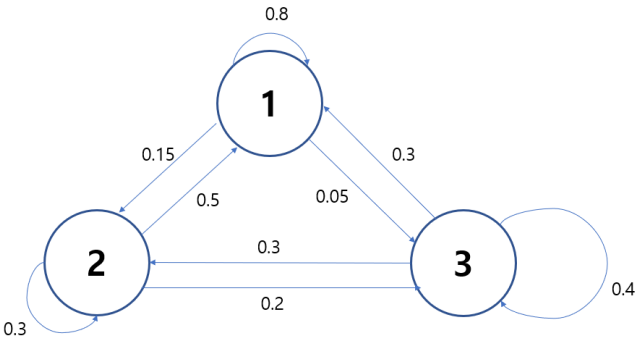


그림 4: None

이성호

Question 1

Suppose that whether or not it rains today depends on previous weather conditions through the last three days. Show how this system may be analyzed by using a Markov chain. How many states are needed?

Question 2

In Problem 1, suppose that if it has rained for the past three days, then it will rain today with probability 0.7; if it did not rain for any of the past three days, then it will rain today with probability 0.3; and in any other case the weather today will, with probability 0.4, be the same as the weather yesterday. Determine P for this Markov chain.

Solution

Let R : Number of cases of rain, U : Number of cases of rain and number of cases of no rain.

The probability of rain or no rain on the next day depends on the results of the previous three days, so the number of possible cases in the last three days is a total of eight: RRR, RRU, RUR, RUU, UUR, and UUUU. Therefore, a total of eight states are required, and the probability of rain in the future is only affected by the previous state, so it can be called a Markov chain.

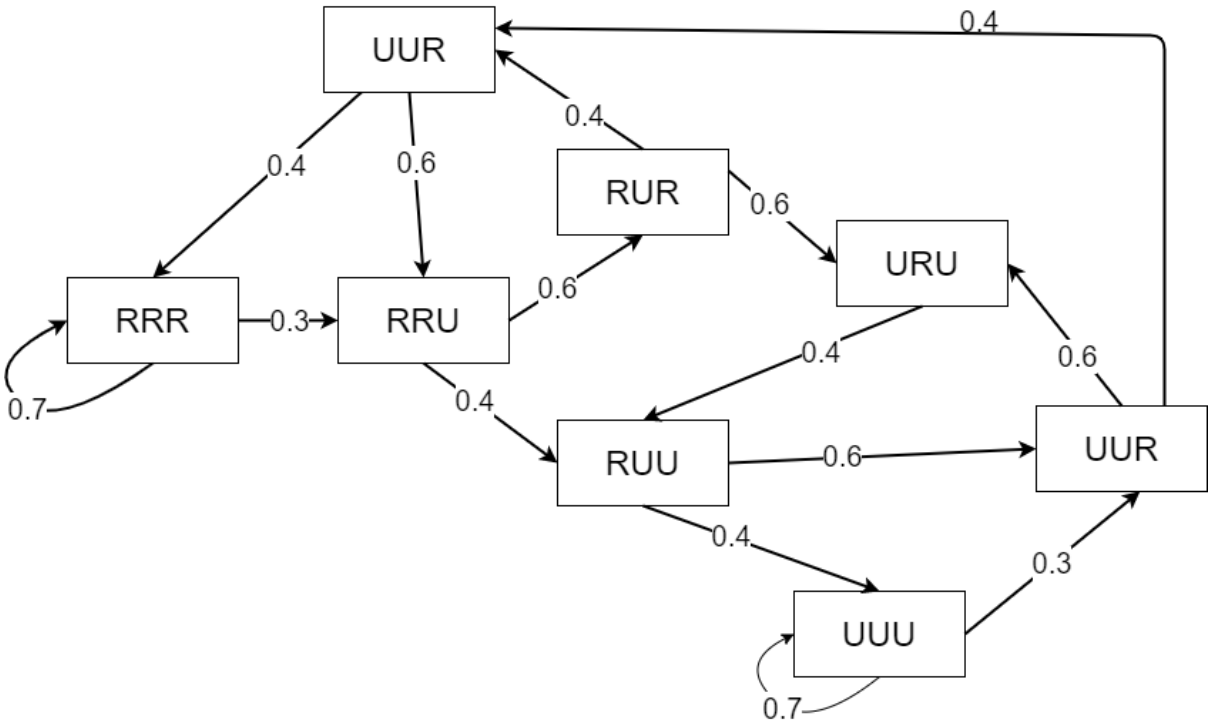


그림 5: Diagram

P		RRR	RRU	RUR	RUU	URR	URU	UUR	UUU
	RRR	0.7	0.3						
	RRU			0.6	0.4				
	RUR					0.4	0.6		
	RUU							0.6	0.4
	URR	0.4	0.6						
	URU			0.6	0.4				
	UUR					0.4	0.6		
	UUU							0.3	0.7

그림 6: Matrix

정원렬 (Effect of Exercise policy for muscle growth)

Introduction

A student who likes to exercise decided to make a strategy for his lower body muscle growth. There are two exercises. First is squat and second is deadlift. If he Squats, he get guaranteed a low injury rate but his muscle grow slowly. However, If he Deadlift, he has a high probability of injury but his muscle grow fast. He have to calculate optimal exercise strategy for maximum muscle growth with optimal effort. As a result, two actions could be defined:

Problem Formulation

- $State = S = \{1, 2, 3, ...\}$
- Action $A = \{Squat, DeadLift\}$
- Squat's Transition Prob $\{0.9, 0.1\}$
- DeadLift's Transition Prob $\{0.7, 0.3\}$
- $Reward \quad R(Squat_{Success}) = 1.5$, $R(Squat_{Failed} = -1), R(DeadLift_{success}), R(DeadLift_{failed})$ 1

Use the Markov chain temperamental difference trait to establish a movement strategy.

Squat Figure

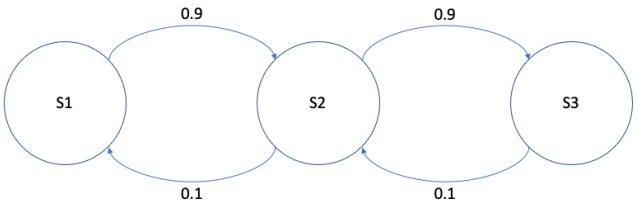


그림 7: Squat

DeadLift Figure

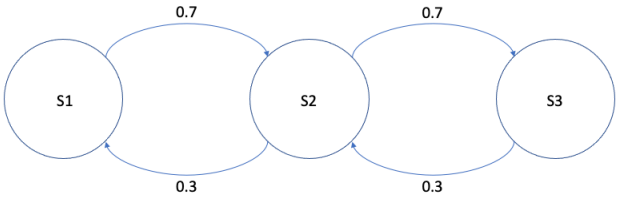


그림 8: Deadlift