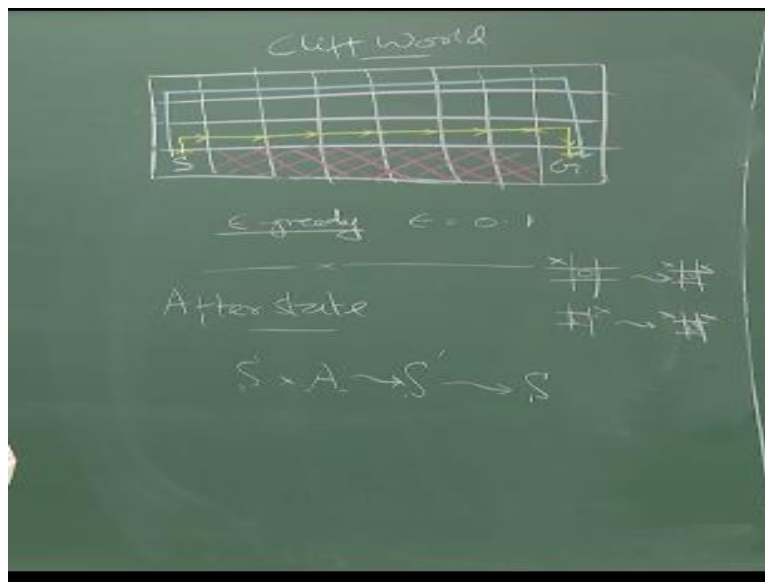


NPTEL
NPTEL ONLINE COURSE
REINFORCEMENT LEARNING

Afterstate

Prof. Balaraman Ravindran
Department of Computer Science and Engineering
Indian Institute of Technology Madras

(Refer Slide Time: 00:16)



But then in the writer I talk about something called after stage I just wanted to mention that very quickly, so what is an after state now it is a very interesting concept yeah what is in after State, yeah so in many problem instances especially things like games or like this kind of a civil mean like a system where that is moves by nature right so there are moves by the human the moves by nature or moves by an opponent right.

So that is the whole stochastic city in the system is introduced either by this opponent or by nature right you are actions the effects of your actions on the system are deterministic right but you do not know what the opponent will do right so in this case of nature is like cows aging and

things like that so your action is selling cows that is deterministic if you send a cow the cow is gone from you right you know that effect right.

You make a move then the piece moves right there is no stochastic there right when I want to move upon on the board I put it there well as long as you do not have Parkinson's or something okay it will actually land up in the square that you wanted to go right and well you have to consider all eventualities right but then you do not know what the opponent is going to do right so if you look at the actual state transition dynamics it is going to look stochastic.

But that is a deterministic component and then there is a stochastic component so sometimes it makes it a lot more convenient to come up with a solution if you pull out the deterministic component separately and represent $s \times a$ as an after state right so what is an after state it is some sense you take your S right you apply A to it so you get to some other space which we will call s' and now all your stochastic is applied to state in s' right and this will result in a in a state S right so this is essentially what is happening so $S \times A$ yet S' , S' to s so if I am going to represent things as s after states instead of learning a Q function I can learn a value over the elements in after States.

I can learn a V function over the entries in s prime this is essentially a combination of $s \times a$ so the v function over s' is almost like learning a Q function and what is the advantage of learning this v function over s' instead of $q / s \times a$, I have one less dimension and anything else one less dimension okay then I explained that in the writer so it could be that many state action paths yield the same s' right.

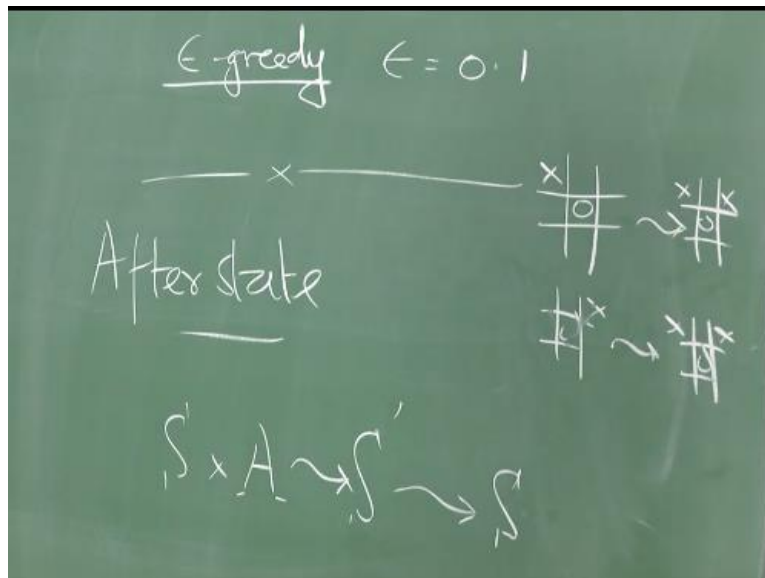
I actually talked about this very briefly talking about Tic Tac toe I told you there are many different ways in which you could have come here right so I do not care which way you got the right I could have for example I could have from there I could go that right offer I could go there right so this is a state and the action was put across there and this is the state the action was put across here but both have the same after state.

And since the evaluation from here is going to be the same but this is an after state right this is a state this is after state right so you see why I denoted the after state by S' why, a tic-tac-toe S is going to have an equal number of X and O's, s' will have one next more than O and if I am playing x whenever I see the board whenever I have to make a move that will be equal number of X and O's right.

But not the state will have one next more than O right so it is actually different set of states they are not really States I mean I might or might not see them in reality for example in the cow task after states are possibly states I mean cannot be stated it they are also states like they are also stays after sets are also states but in detect O case after states are not steps because you will never see a state that has a more access than O's.

So that is why that is why I denote it this is s' as a separate thing it could be the same as S but it could also be different but the point here was this is actually a QSA right I mean if I learn a value function over the states its QSA and since the whole thing is deterministic.

(Refer Slide Time: 05:41)



If I know the rules of the game I can do the look-ahead so I give you a state I look at all possible after States and then I pick an action which leads me to the best possible after state, now I can do the look ahead because it is only the application of the action I do not have to model nature and I only have to model what happens when I apply my action to it so it is a possible look ahead and this is how Gerry Tesauro implemented is backgammon also.

Now he did not learn a value function over state axes you learn to value function over after States right back and also write once the die roll is determined that you take an action right when you take an action your state includes the die roll and once you know the die roll then whatever action you do the effects will be deterministic and then they are opponent will actually make a non-deterministic move so that is fine, okay any questions?

Okay so that means is the end of chapter 6 so at some point I will do some proof sketches for TD 0 and possibly q-learning but it might not be next week we will see and so next I will start off on Chapter seven.

IIT Madras Production

Funded by

Department of Higher Education

Ministry of Human Resource Development

Government of India

www.nptel.ac.in

Copyrights Reserved