NPTEL

NPTEL ONLINE COURSE

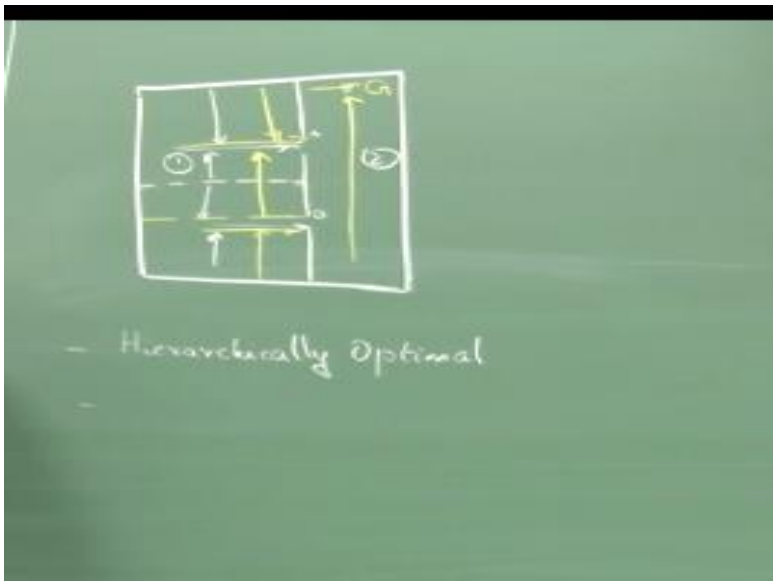
REINFORCEMENT LEARNING

Types of Optimality

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

So there are two things you say need wanted to do today okay let me do the easier one, right.

(Refer Slide Time: 00:30)



So I have a grid world or friendly grid well I have not drawn the grids okay I assume that is a grid here and I want to go from room 1 to room 2 room okay there are two doorways which is a

best way for me to know no such state we could start from anywhere in the world what is the best way to go to room 1 sorry room 2, sorry depends on what I am asking for policies I mean what do you mean it depends on the start state give me a policy in the room 1 when to go to room 2.

Does not matter what start status what distance pick the closest store okay great so what should I do let us say split this difference here right, so if I am below this I will go down if I am on this line I will go right if I am below this I will go up right, so this is my policy right likewise I will go up here if I am on this line and all right remember this I will go to right so this is my policy all of you agree with this great.

Now If I tell you that my final goal is here it is a policy the same one as a change, how many of you say it is the same, yeah I mean the standard reward I mean minus one for every time step and G when you reach the goal you get 0 or 1 if you reach the goal and gamma discount factor does not matter standard shortest path rewards one of the two shadows particles know any step you take minus one if you need to know the reward yeah.

I am saying it is just the standard shortest path reward the answer to my question does not record depend on the work function it could be minus one for every state it could be one with the discount factor of gamma right one at the goal state and zero everywhere else and gamma is pointing in something does not matter sure it is yeah, how many of you say that answer would not change?

Well I mean if you want to know the policy in room 2, it is okay that is a policy in room two so anywhere below this line you go up that line you go right okay so that does not matter that does not change right now we need to ask so what is the policy in room, how many of you say it will change? Because people who did not put their hands up for both questions what is your opinion, you did not you could not.

I could not raise my hand I thought you had some problems with your shoulder or something okay it would not change would not change no one nobody gave the right answer, so if you think

about the question of solving the problem right solving the problem of going to the goal state this is actually right so it is basically that right is not matter I mean I can either go right or I can continue going up.

So this right or up is this is a matter of choice in this line but I just chose to go right okay it shows now there are two policies here the white policy for room1 and a yellow policy for room which one is correct which one is optimal let me not say correct, now which one is optimal, yellow is optimal I will say both yellow and white are optimal, depends on how we define an optimality right.

So when you talk about this kind of hierarchical problems so you have different notions of optimality right so you have something called hierarchically optimal right so what do you mean by hierarchically optimal, so a solution is hierarchically optimal if I say that you have to search through policies okay you have to search through policies that respect this hierarchy I will give you some hierarchy right.

So in this case what is the hierarchy I want to solve any problem of going to room 2 going anywhere in room 2 I first use the sub problem of going for out of room 1and I go out of rumen and then I go to room 2 this is the structure right so I have to go out of room 1 and then they go to obviously that is only structure you can have here but in general I am saying suppose I want to go to the I do not know say you want to go to go to your mess.

Right and I say that the only way you can do this is first you have to go to DC right and first you have to go to GC and then from GC you can go to the mess there might be other shortcuts which you can get by going through the admin block or something like that right which I am prohibiting now right, so there might be an optimal solution that takes you to miss without going through GC.

But I have given you a hierarchical structure right like you gave me a hierarchical it how do you get to the gate he said go to the bust and get a bus and then go right I could have come to drive I could have actually gotten somebody to give me a ride out outside the ICS are building you have

prohibited that policy right sorry yeah whatever so you pure prohibited that policy or told me go get a bus here right.

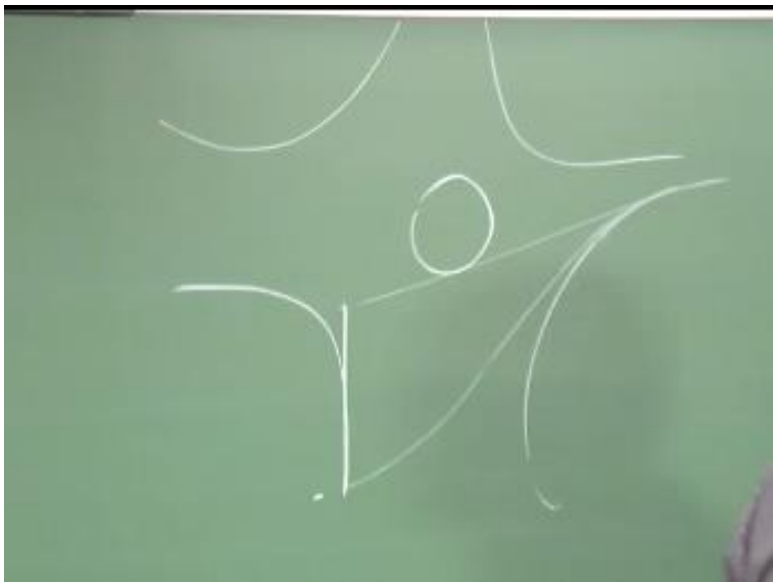
So that is the what I mean by hierarchical structure so consistent with the hierarchy that have given you what is the best that you can do right notice that I am NOT asking you what is the best way I can solve each of the sub problem what is the overall best I can do, right so going back to hear right so what is the best way I can solve the problem of getting out of room one just without any reference to anything else.

What is the optimal the best simple way of getting out of room at the White policy right, so if I am hierarchically consistent with getting out of room one and then go to the thing and I could potentially have used the white policy and then use the yellow policy in two okay that is a valid solution okay but overall among all policies that satisfy the hierarchy that is not the optimal one in terms of overall cost that I am going to get.

I say doing the yellow policy in room 1 followed by the yellow policy in room 2 gives me something that is hierarchically optimal okay following the white policy in room one gives me a solution where the individual components in the hierarchy are optimal let the individual components are optimal but overall put together that is not the best possible solution right, so for example I have told you go to GC and then go to the mess right.

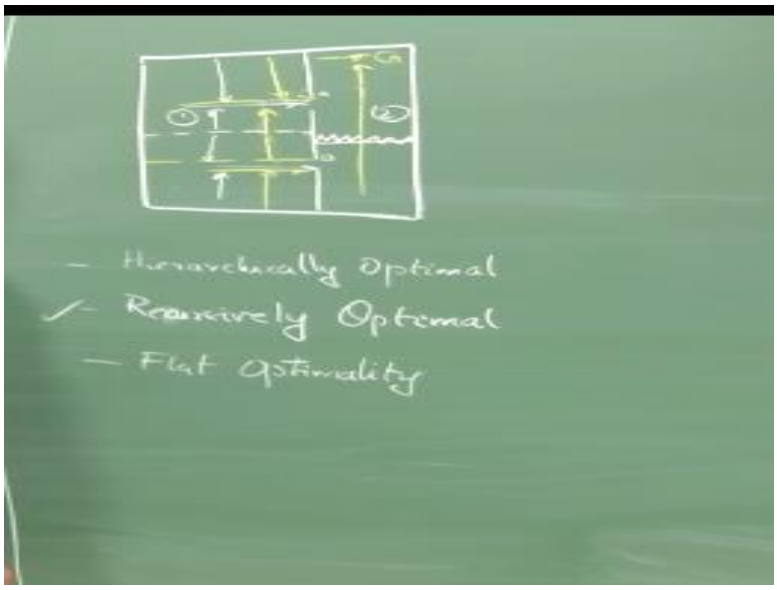
If I know that going to the mess is my is my ultimate ambition right I might try to cross the roll over in front of ICS are and then walk in some kind of a diagonal barely satisfying the definition of going to GC and then turn right, so I am going from here HTC.

(Refer Slide Time: 09:40)



So from ICS are I say go to mess but go through GC so you could possibly do something like that right that is hierarchically optimal but if I say go to GC first optimally so then you probably do this right and then from here now that you have reached GC now go to the mess then you will probably do this so obviously this is not hierarchically it is not optimal right because that is a solution which satisfies going to GC and going to the mess in a shorter time overall shorter time than this one that I have solved right. So this is essentially solving individual components optimally right.

(Refer Slide Time: 10:24)



so if I come up with a solution where the individual components of salt optimally right and then the overall solution is subtend by putting the individual components to get the right then I call that I call that recursively optimal right, so given that all my sub problems are optimal and the policies are frozen what is the best can I do is I have to put together this optimal sub policies to find something that is best among those policies so this is a more restricted policy right so the hierarchical optimal or did I say I have to respect the hierarchy but I need not necessarily solve each component optimally.

Overall the problem should be optimally solved what is recursively optimal say he not only should I respect the hierarchy but for each individual components in the sub hierarchy I should have an optimal solution so given that everybody under me is optimal what is the best solution I can get okay so and the third which sometimes called flat optimality is essentially this solution you are talking about right.

So I cross I say sir get to the admin block and then find away to get out there right so that does not respect any of the hierarchical restrictions right it just looks at the most basic actions that are possible in this world and tries to find what is optimal okay, so usually flat optimality gives you a more quote and quote more optimal solution than hierarchical optimality which in turn gives you something better than recursive optimality right.

Overall cost wise correct so which one should you which one should we look at which is this is a better solution to have, how are you guys computational program okay anything else flat optimality essentially I mean I mean whole from the beginning I have been telling you why hierarchies are useful there are so much is advantageous to that and so on so forth for finding flat optimal policies.

You will have to give up on hierarchy right so yeah so if you have enough computation power and everything you can just ignore hierarchies and then solve everything from scratch right if you have enough computation power you do not even have to use function approximation is it correct, think about it anyway so yeah so what flat optimality is hard to achieve so we are given up on that so now the question is between hierarchical optimality and recursive optimality which one do you want to achieve.

So people understand the difference between the three right okay so I can ask me questions on that in the exams what will be flat optimal in which case this problem this case that optimality is the same as taking the yellow policy in this case it is the same there is no difference that is why I do the agenda a circle picture so that you know that is flat optimal that is hierarchically optimal and that is recursively optimal that optimal is globally optimal.

Now this your position I am still waiting for an answer there was a question hanging hierarchical optimal or recursive optimal which is better, recursive is better why, recursive is better because it is easier to transfer and the anything else for the other advantage is its faster because you have to sorry yeah I leave I need to remember this is recursively optimal I still have to solve the whole problem not that I am not solving the whole problem right.

So but I can modularize it as you are saying right so I the learning horizon is a lot simpler in terms of solving recursively optimal problems provided I have a way of learning to solve the entire problem by treating this solutions to these sub problems as actions themselves right so I need to have that infrastructure in place if I have that structure in place then I can go ahead and solve the sub problems first and then put the solutions together if I do not have such a structure in place then it becomes tricky.

I mean I can still solve the sub problems but still I will have to look at the whole problem at the higher level right so that makes it real hard so I so recasting optimal is great because it gels with many of the things you are talking about right so it allows you to transfer reusability in the things like that so why is it more reusable because I am learning something had a hierarchically optimal you saw that I learnt a weird policy to get to the circle.

So that is not how I would like to get to the gauge in the circle every time right so I will really like to walk straight most more often than not using the shortest path together the very end result will probably be more might not be the best for any particular sub problem not sub problem for any particular problem it might not be the best way but across different place of different problems for which I can use the agenda a circle as a stopping point maybe this will be the better one on an average.

So it is always a good idea to solve these things independently so that I have to worry about my TD error from solving the original problem propagating back to solving the sub problem they think of how I am going to learn if I have to learn the yellow policy that essentially means that the reward that I get here okay has to propagate all the way back to states here right only then this will start getting updated to go out like that right.

So it becomes a little trickier learning problem so longer horizon and stuff like that right but when do you want to learn hierarchically optimal policies, now I am solving the same problem multiple times I really want to get a solution for that or my recursive optimal policy is really bad it is possible very well possible that there is this corner case where the optimal sub solution is really bad right.

So I would really like to get the hierarchical for example right I put a laser beam right so every time you get out like this and it rode to go up there you get burned by the lesser ones ironically still the white policies recursively of okay but he do not want the white policy he want the yellow here, Oh really there is no spectrum here right it would be like these are saying come to the upper half but it is still not still you are making it hierarchically I mean recursively optimal.

But with more options but I keep feeding it to the point of hierarchy happening no he is still having the hierarchy right he is not still allowing it to go through the admin building he is still saying you have to use one of these ways of getting to the Agenda circle and then he can solve it right so in some sense what you are essentially doing is you are increasing the branching factor of your hierarchy.

So yeah I mean sure but all that is happening is it still says recursively optimal till your options reach a continuum where it becomes hierarchical optimal but it is not like there is a continuum of hierarchical and recursive essentially there is a different qualities of policies that you can get there are still recursively optimal and in the limit becoming higher is same as a hierarchical optimal case there is no not necessarily continuum of characterizations of things that are recursively versus hierarchical yeah.

So I agree with what you say we need so but then why would you want to define so many hierarchies so many so many sub problems anybody who can see but at the same time we create so typically we would do that if it warrants if it is warranted by the problem settings for example like this here right now there is a lesser there and maybe making another door somewhere there and hey exiting through that is a safer option right.

so we can do this but the point I wanted to make is even with that laser thing that the white is recursively optimal right and you do not want the recursive optimal policy in this case because the value of the optimal policy is really bad so you want to do something that is hierarchically optimal in this case because here the new of the laser on the flat optimal solution will be much better then what we had.

Because you actually exit from the region two the lower bottom you will exit through the door and then come back to the region that would be the flat of your enemy okay, so if I have any restrictions on the on room two as to what you have to follow if I say room two this is the only action allowed to you yeah the hierarchy comfortable policy will be bad and if I in room two I can say okay you can go you have an option of going to either room one or going to the goal then it becomes okay.

Mean depends on how I defend the hierarchy rock is pointing right but if the hierarchy says only action here is go to room two okay then it is a bad option yeah you are right okay, so now people wanted a difference of what is the difference between flat optimal hierarchical optimal in this world right I said it is the same but once you put the laser it is no longer the same right so because in room two I have only the yellow action here right.

So if I am not in room 1 if I am in the bottom half of room to I am going to run into the laser and the top of room to I am fine but a flat optimal would be if you are in the bottom half go to room 1and then go back to room towards the top of that will be the flat optimal case okay, so i can give you wells like this and asked you what is recursively optimal what is hierarchically optimal and so on so forth.

IIT Madras Production

Funded by

Department of Higher Education

Ministry of Human Resource Development

Government of India

www.nptel.ac.in