Reinforcement Learning Interview – Sutton – 2016.11.16

1. **From behaviorism to reinforcement learning**
    a. **You received your B.A. degree in Psychology, then you turned to computer science. What is the reason for the direction-changing?**
        i. It's not a direction change; I was interested in how learning works as most of psychologists concerned about it, and I got Psychology degree in 1977; at that time learning is not popular in CS; I always want to study AI, then starting CS from MS, PhD. My views of AI is recolored by Psychology with human and animal learning(which is my secret weapon), because many people in AI don't have this background, I started there and got lots of inspiration from Psychology.
    b. **Do you believe in Behaviorism? Inspired by Behaviorism?**
        i. I was inspired by Behaviorism. But never understood why it has such a bad reputation. I just took useful things from it.
2. **History of reinforcement learning**
    a. **How reinforcement learning started? starting point to write algorithms?**
        i. In 70s there is machine learning and is popular but no such thing as reinforcement learning. To me it was always a obvious idea, a learning system wants sth and bear some behaviors to achieve sth and it's not present in ML.
        ii. Some one called Harry Klopf, he wrote several reports, pointing out that this kind of learning is missing. We don't have learning agent tries to achieve sth.
        iii. We tried to figure out what the basic idea, and found out that he is right. A system wants sth and bear its behavior to achieve it. This idea has never been studied in any fields, especially in ML, not Control Theory, not in Engineering, not in Pattern Recognition. All those fields overlooked this idea. You could see some earlier work in 50s, people talked about trialneuro(6:00) but in the end it becomes supervised learning. SL which deep learning is, they have targets and training sets and try to memorize, try to generalize from it. It's funny we're nowadays talking about DL and RL. Way back beginning, it was the same issue, try to distinguish RL form SL. I said there is sth new. And it's different any more. Still, people overlook the need for sth more than just supervised training. What you need more, that you need a system that can learn and that's all! So RL systems finds a way behaving or maximizing the wolrd, where SL just memorize the example given to them, generalize new ones but they have to be told what to do. RL systems can try different things. We must try different things, we must search actions and spaces or define learning to maximize the world. So that idea has been lost and Andrew Barto(mentor) and I gradually realize that it's not present in old works and it was needed. This is simplified view of why we're precursors. Eventually we tried to do basics.
    b. **Major achievement of U of A RLAI**
        i. **How RL AI started - GTE, AT& T, 200. It has been a very long time since you have been developing and promoting Reinforcement Learning (RL)**

from 1979. **When you started research in RL, the computing power of the computers at that time was very limited, so it must be quite difficult to apply RL to a real-world problem, as Reinforcement Learning involves a lot of trail-and-errors before converging to the optimal policy, which can take an extremely long time. What gave you faith for RL at that time?**

1. does not agree with that the mean of RL is slow, i don't accept that. But I do accept that the fact that increasing computational resources have a big impact on this field. You have time to coincide with the availability of hardware. We have to scale expectations when the availability of hardware. Still a bit early for deep learning, deep learning successfully use a lot of computation because its strength. RL could be viewed as similar or dissimilar, but really the larger thing you want to do in RL require great computation. It's been a long time now since 1996 people said we'll have computation power for strong AI in 2030. I think it is not just only depend on cheap hardware, but also algorithms. For example, if we have algorithms by then, I don't think we have algorithms now but we might have algorithms by 2030.
2. It's a big question whether one of the hardware first or the software first. We have the software to test out hardware, and the availability of hardware pushes people to software. But it's not tremendously valuable for smallest guy researching or working in limited computational resources. Even in 2030 we may have adequate hardware, we may still need 10 years more for smallest guy to catch up with algorithms. Now you know my reasonings, you can reevaluate it or change it yourself.

ii. **Why did you name your lab Reinforcement Learning and Artificial Intelligence? Is RL a solution to all AI problems? Can you give more explanation?**

1. Some people think RL is just Reinforcement of AI problems, actually, Reinforcement learning problem is an abstracted approach to AI. I'd like to say we're using an approach to AI. It's funny to name Reinforcement Learning and Artificial Intelligence, the word 'and' in English can mean either exclusive or inclusive, it can be and or can be or. Because Reinforcement Learning is both a subset of AI, and also originate of AI. It's quiet ambiguous. We're still looking for an answer.

iii. ~~**Follow up: With recent stunning achievements like DQN and AlphaGo, do we have more confidence? Is deep reinforcement learning the solution to AI? What do we still miss in achieving strong/true AI ?**~~

c. **Applications**

i. **For people who are interested in reinforcement learning ?**

ii. ~~**How do you see the future of Deep Reinforcement Learning in 10 years ?**~~

iii. ~~**What's the benefit of using RL to learn knowledge? Can we transfer knowledge from the policy of one problem learned by RL to another problem ?**~~

iv. **in CV(Computer Vision), some one says that if a CV lab is not working on DRL, its not promising, how do you think of this claim?**

1. It is correct that deep learning and reinforcement learning
2. Combination of RL and DL is a really good combination. That's a good combination. And you can certainly do a computer vision without reinforcement learning and practice normally to do it as how to prepare dataset mainly supervised example and then to learn from that. But I can say you couldn't have it without deep learning. But who would actually take some imaginations and do it with reinforcement learning. I think it would take some cleverness and imagination to do that. I'd tend to think that would be a breakthrough to do computer vision with a degree of reinforcement.
3. The winning feature of reinforcement learning is that you can learn during normal operation. Conventional deep learning learns from trained label training to do so that means that once it's in place or once the phone (it is) out in the world doesn't learn anymore. (RL) Whereas in principle you could learn from your normal operation. You could take some imagination to reform it because you don't have examples but you have much more experience than just normal use. And then you do(test) in the training examples.

v. ~~**What do you expect about practical applications of RL, in particular deep RL, in areas like robotics, health, operations research, etc? Is any area that you do not see RL could fit in?**~~

vi. ~~**Any area**~~

d. **Future of RL**

i. **Challenges**

1. **What are fundamental research issues in RL?**
2. Technical ones
3. Limitation of reinforcement learning and AI in general?
4. Can you plan the learned problem with the world
5. what you mean by knowledge
6. what if behavior are formulated in different way
7. Off-policy learning is biggest challenge with functional approximation
8. Well there are several really important ones. There are technical ones. But let me go towards something that we can all understand, which has the harder limitations. Reinforcement the AI in general, which is we would like to be able to learn how the world works and then apply that knowledge in our plan, corrects autonomy behavior. So we take something like AlphaGo or computer chess. We don't have to learn how the world works. We know what the moves are and we know what the consequences of moves are or we move this piece there then the board will be. And you know we can already do amazing things in term planning scenario like that. We like to do the same thing we have the moves the actions the choices and the consequences are learned. We got a new mechanism, new plan with the learned model of the world. That's the key problem I think. We have no choices

and no consequences to make models of how the dynamic of the world states and demonstrates. **Once we got that sense, we will be able to plan them and to do AI in stronger sense.** There are subproblems what mean by knowledge. What kind of predictions we want to make about what will happen in different ways, how we formally behave in different ways. That technical problem is my favorite type, of this. What we have different ways of behaving or what we have when there are consequences. We're going to learn the consequences from trying out different ways but **without taking them to completion** just the normal thing you walk into the room you say OK Right here water I have a chair. I have different objects people that I can talk to all these are opportunity reaction now but I will only do one thing and maybe I will never pick up the bottled water drinker but I learned from looking at it what it is. **and something you learn from these partial experiences which we call off-policy learning so off-policy learning is our big technical challenge in reinforcement learning.** To learn efficiently off-policy function approximation. you want to learn in a scale way, you want to take unprepared data, you don't have to have a training set always label pictures, you want just be able to interact with the world and gain experience and learn the way the world works from them so how can we learn from unprepared experience with world. **That's what reinforcement methods are good for and should be good for.**

9. ~~**Google is using RL to train robotics, is it a sign ? What are the bottleneck and the limitation of Reinforcement Learning? Credit assignment problem or exploration-exploitation dilemma ?**~~

ii. **RL + DL = GL?**

1. ~~**Nowadays, Deep Reinforcement Learning which combines Reinforcement Learning with Deep Learning has become a very popular approach to solve many kinds of problems, such as game playing, decision problems, robotic control etc. Did you expect this 20 years ago?**~~
2. ~~**Deep Q learning used a deep neural network to approximate the Q function in Q-learning instead of using a look-up table or linear equations (Traditional methods). Will combining deep learning with RL be a general promising way for achieve better results? Why or Why not?**~~
3. ~~**Do you think Deep Reinforcement Learning is the most promising approach for developing Artificial Intelligence for the robot? If not, then what do you think that should be?**~~
4. **AI benefit much from psychology and neuroscience, like RL itself and ConvNets. You add two new chapters in your new edition of RL book. What are important interactions between AI/RL and psychology and neuroscience?**
5. You readers may not know but the basic reinforcement that trumps different learning has been **essentially found in the brain**. There are processes in the brain that look like they're following the same rules that are well modeled by the rules of reinforcement learning to put difference learning and that's been known for quite a while now. And it is fact that the

standard model of the world systems in the brain. **Is the reinforcement learning type of difference model?** And I say **standard model** because it's not that (it's) perfect but as the standard model everyone picks on it so you know you succeeded when everyone this is picking at you. **Your reward systems in the brain is also pretty good model of animal learning psychology behavioral learning.** The other major thing is a model based learning where you can do planning. The planning and the brain are all based on various notions of replay imagining circumstances. And like there are lots of demonstrations that like rats were going to match and taking paths are a maze. That's also a model reinforced how we plan, we imagine sequences that we learn from that was there and so those are all major things that's interesting. The thing to think of both great researchers in AI researchers trying to figure out the mind and it's reassuring sometimes because if one of them fails maybe others succeed.

3. **Thoughts on Topics (Gather from our readers)**
    a. ~~**Google DeepMind has developed many state-of-art research on Deep Reinforcement Learning, especially the recent advancement in AlphaGo that defeats Lee Sedol, one of the best Go player in the world. One of the hero behind the AlphaGo, David Silver, was your student. How do you think about the research of Google DeepMind team?**~~
    b. **Why self-play is so important, and works well, e,g,, in AlphaGo? Is there a limit for self-play? Can an agent keeps improving its performance?**
        i. can generate infinite training data
        ii. limitation → its good when we know all the rule of the game
        iii. limitation is that you need just to play with yourself, you have to be able to generate the risks of the consequences of those moves works great for software games again if you know the consequences your moves you can play them out in your head. **The limitation is in regular life, we don't have an analogous to the rules of the game, just tells us how good the pieces of your real life.** You know you pick up the phone you press a button or something will happen. **You have to learn that you don't have the rules of the game built-in. You don't know the consequences of the moves.** So did the self play you need the rules of the game.
        iv. The big strength of South play is that you can use it to **generate an infinite amount of training data** you don't have to have people labeling the training data to play yourself if you can number the examples, different games. **That's what we want.** So can we do something like self-play for real life not just a game.
4. 1961(Minsky)
    a. **from Video and Paper: A Batch, Off-Policy Actor–Critic Algorithm for Optimizing the Average Reward**
        i. **We watched the video on Youtube which you gave a speech about 'the future of AI' in February. You said that "Moore's law" has a large impact**

on deep learning while the core methods (back propagation, Neural networks(NN)) has already made during 80s-90s. You give a example: "NN" won because their performance scale with 'Moore's law' and the best algorithm is just as same as 80s.

1. this is a good really the best example of the need for our rooms to be scalable and the importance of Moore's Law. Moore's Law was just shorthand for this mega trend of increasing competition for decreasing dollars. In the 80s, you know neural network is a big thing, but they met their limitations and became a little bit passe(out of date?) there just in the great waned. And now essentially the same methods are extremely popular again. Fifteen twenty years of Moore's Law just made the scalable method much better.
2. What's the alternative being scale. You can get good performance because you have a method that scales for competition and so a few more competition give it the better work or you can put in human knowledge. You could say I have an idea how we see or I have no idea how we process speech and and that using human knowledge is always going to make things work a little bit better but it doesn't scale at all. The way ten years human knowledge is still just you put more work into it it will be form a little bit better. But if you scale right, then you just have to let time go by and your methods will get better and better and better and I think that's the lesson that we've seen in neural networks for for speech and generally for Image recognition and computer games all over the place.
3. So it's unpopular because if you're a researcher, what you want to do is to provide something as a person or just a person you want to do something for that. What you want to provide your personal input that makes your system work. if the Personally input is that your system needs to work well is just, you know, to wait five more years.

   It's kind of unsatisfying. I mean it's you don't feel that you're personally responsible and so people always want to believe that their input is making things better and it's unsatisfying. But that's what we should do better there be some sense.

ii. **Could you please talk about the importance of scalability for deep learning**

iii. **Can we say that the current development of AI algorithms and methods is slower than the development of hardware?**

iv. ~~**Nowadays, more and more scientists are making effort to develop distributed deep learning methods which can greatly improve the speed in a direct way. What do you think about CNTK (Computational network toolkit) which is scalable open-source deep-learning toolkit developed by Microsoft. Will it save researchers' time on deployment and have more time focus on the improvement of algorithm?**~~

v. ~~**Is current reinforcement learning scalable?**~~

vi. ~~**Deep reinforcement learning is very good at some games like Atari and Go.**~~

~~How about games like WarCraft?~~

b. **Deep learning hungers for big data. Reinforcement learning usually also needs lots of samples. However, there is research on one shot learning, trying to learn with one or a few samples. This may be the way people learn for some problems. Is it possible to integrate the idea of one shot learning with RL?**

    i. No reason it has to be slow. It has to come up with representations in advance. I once said that everyone knows we try to find good representation so that we can learn fast and we're just making sure I want to check if you guys, are you are you excited about deep learning because you're able to find good representations so they can learn fast and they said no! They were excited because they were able to learn complicated functions but they accepted that would be slow to learn. But it's still true that the problem of slow is still unsolved.

    ii. Learning slow so that you can learn fast; learning from one shot. I have this phrase you have to learnings learning slow so that you can learn fast. So you know people through our lives we learn good representations. So that then when we get some experience we can learn very quickly what the correct behaviors mean we can learn from one shot but that learning from one shot builds on a long period of gathering representations.

c. **As a well-respected researcher and leader in CS, how do you think the dominating status of electric magics nowadays?**

    i. So for me it's really science, it's trying to understand the mind. That's the first goal and I spew if we understand it, it will be to have useful applications and spinoffs.

d. **Can you give some advice to the beginners of RL? How to study and which part we need more researchers to work on? Application wise and theory wise?**

    i. Learn the basics

    ii. Find application that make decisions that aren't high cost

    iii. Example: there's a known correct response on something where the the correct response can be deduced from the data. So for example here's a decision that I made. Says, think every day I'm on an escalator. Sometimes it's better for it to stop because there's no one in the middle of night, maybe there's no one coming out yesterday. And so you want to save energy by shutting down and but then when the person arrives you want to turn it on. So how do you make a schedule for turning it on and off so that it's running most of the time that a person arrives and it shutdown if no one's come for a while. And so is the information available and the data information is available dated when the new person arrives and you're not running. That's a bad event when you're running the elevator and no one is coming. You're just wasting electricity that's that's also miserable that. So you can use that in such a case you can use the actual data without training information to do so. Think of things like that anywhere this is.

1. ~~Should read the book~~
    a. ~~Based on behaviorism~~