Artificial Intelligence Coursework

Waseem Syed

1 Reinforcement Learning

Sutton & Barto (2011) says that the term 'reinforcement learning' is used to denote a more general approach, so that an action applied to the environment in a particular state produces an immediate reward and a change of state. The long-term return from the action is sum of the immediate reward and the return achievable from the new state. The return obtainable from the each state, or 'value' of the state, is therefore evaluated by a form of bootstrapping, since the value of any one state depends on the values of others that can be reached from it. This is an approach of a system, which learns from experience. Most work on neural nets is directed at "learning with a teacher". Reinforcement learning is focused on interaction between the learning agent and unknown environment.

In Lecture Notes al Rifae (2019) explained that interaction with environment is a major source of knowledge. Reinforcement Learning is much more focused on goal-directed learning form interaction than are other approaches to machine learning and it uses aspects of computer science, psychology, neuroscience, engineering, economics and mathematics. There are few basic elements in Reinforcement Learning which are, a policy, a reward signal, a value function and optionally a model. A policy is a kind of mapping from perceived states of environment to actions to be taken when in those states. A reward is a scalar feedback signal which measures how well an agent is functioning at a given time and the agent has task to get maximum rewards. Reinforcement learning is based on reward hypothesis. The value function illustrates what is good in the long run. It is the total amount of reward a state is expected to accumulate, starting from that state, over the future. A state can have very low immediate reward but a high value. There is a difference between values and rewards. Values are predictions of rewards and there are no values without rewards and values are estimated and re-estimated from a sequence of agent observations over its lifetime, on the other hand rewards are directly determined by the environment.

Jaderberg et al. (2016) states that theory of reinforcement learning is rooted in psychology and neuroscientific perspectives on animal behaviour, of how agents can improve their control of an environment. Humans and animals solve problems through a combination of reinforcement learning and hierarchical sensory processing systems. On the other hand reinforcement learning agents have also achieved some successes in a variety of domains.

Wiering & Van Otterlo (2012) tells that the goal in reinforcement learning is to develop efficient learning algorithms and it is of great interest because of the large number of practical applications, from problems in AI to operations research or control engineering.

Dayan & Balleine (2002) discussed the engineering theory of optimal control and the computer science theory of reinforcement learning, both study how systems of any sort can choose their actions to maximize rewards or minimize punishments. They evaluated the actor-critic model of dopamine system from the perspective of the substantial psychological and neurobiological data on motivation, which is key to the modern view of reward learning. Reinforcement learning comprises one family of methods for performing dynamic programming, using the values of states as a form of cache for the reward consequences of the actions subsequent to those states.

1.1 Conclusion and Evaluation

The evolution of different algorithms with respect to the environment complexity would be one key feature to examine and a better indicator of progress in reinforcement learning. Temporal difference (TD) learning is a concept central to reinforcement learning, in which learning happens through the iterative correction of your estimated returns towards a more accurate target return. There are three most in influential algorithms, Temporal Difference (TD) Learning , adaptive Actor-Critics and Q-learning in RL. Naturally, the application to the evaluation of Reinforcement Learning systems is much more straightforward than other AI systems, since the test framework and reinforcement learning are based on the notion of interacting with an environment through observations, actions and rewards But the most important and difficult part is trade-off between exploration and exploitation which is a challenge in reinforcement learning. It could be said that there are two fundamental difficulties one encounters while solving reinforcement learning problems: the balance of exploration vs. exploitation and long term credit assignment. The question that agents are unable to answer, do I keep following this policy that's giving me nice rewards, or do I take some relatively suboptimal actions now in case there's a possibly bigger pay-off later? This problem is so hard because there can be no right answer in general - there is always a trade-off.

There is, of course, much work ahead. One clear area for future work is the evaluation of other reinforcement learning algorithms and the analysis of the parameters in all algorithms. Finally,tests for humans and (non-human) animals will also be a very important source of information to see whether this top-down approach for measuring performance and intelligence can become mainstream in

AI.

Despite some challenging problems in Reinforcement Learning, I still believe it is the best framework we have right now to work on general intelligence. Otherwise. Watching DQN play atari from visual input, or AlphaGo defeating the world champion in Go, x Flip pancake, x Autonomous drifting, x fly autonomous stunt helicopter, are truly impressive moments in which we witnessed an advancement in Reinforcement learning.

References

al Rifae, D. M. (2019), ‘Lecture notes’.

Dayan, P. & Balleine, B. W. (2002), ‘Reward, motivation, and reinforcement learning’, *Neuron* **36**(2), 285–298.

Jaderberg, M., Mnih, V., Czarnecki, W. M., Schaul, T., Leibo, J. Z., Silver, D. & Kavukcuoglu, K. (2016), ‘Reinforcement learning with unsupervised auxiliary tasks’, *arXiv preprint arXiv:1611.05397* .

Sutton, R. S. & Barto, A. G. (2011), ‘Reinforcement learning: An introduction’.

Wiering, M. & Van Otterlo, M. (2012), ‘Reinforcement learning’, *Adaptation, learning, and optimization* **12**, 3.

2 Question

Write a research essay/report on temporal logics and/or their application in the domain of Artificial Intelligence. It can be about any relevant theoretical and / or practical topics, such as:

- Time Theories and / or Models
- Temporal Knowledge Representation and Management
- Temporal Data Mining or Case-Based Reasoning
- Time Series and / or State Sequences
- Temporal Database Management
- Reasoning about action, event and change
- Prediction / Planning
- Diagnosis / Explanation
- Industrial Process Control
- Historical Reconstruction
- Natural Language Understanding

etc., but only focusing on a SINGLE topic. You may make use of materials which you find in the lecture notes, textbooks and the Internet, but you should adapt them to your essay and give full citations and references to sources each time copied material is used. The contents of the essay/report must be related to the key words "Time" and/or "temporal", in terms of a well-presented literature review/survey (8 marks), together with your own understanding, observations, critical analysis and evaluation of temporal logics and/or their applications in the domain of Artificial Intelligence (12 marks). The essay/report should be around 5 pages using Times New Roman Font, 12 point, with 1.5 spacing, including references and Web/Book citations. Longer submissions will not be penalized but will not necessarily draw extra credit.20 marks

3 Question

Assume that the universe of discourse is the set of people studying or working at the University of Greenwich. Rewrite the following statements in the form of predicate logic.

3.1 (a)

Each person is either a student or a staff.

$\forall$x(student(x)$\vee$staff(x))

3.2 (b)

Each lecturer teaches some modules.

$\forall$ x $\exists$y(lecturer(x)$\rightarrow$teaches(x,y)$\wedge$ module(y))

3.3 (c)

Some hard-working people are not boring.

$\exists$x(hard-working people(x) $\wedge$ $\sim$boring(x))

3.4 (d)

Hard-working people are respectable.

$\forall$ x(hard-working people(x) $\rightarrow$ respectable(x)

3.5 (e)

Everyone knows some hard-working people.

$\forall x\ \exists y(\text{hard-working}(y) \land \text{knows}(x,y)))$

4 Question

4.1 (a)

Use a truth table to verify the following equivalence:

$A \land B \to C \equiv \sim A \lor \sim B \lor C$

A	B	C	$\sim A$	$\sim B$	$A \land B$	$\sim A \lor \sim B$	$A \land B \to C$	$\sim A \lor \sim B \lor C$
T	T	T	F	F	T	F	T	T
T	T	F	F	F	T	F	F	F
T	F	T	F	T	F	T	T	T
T	F	F	F	T	F	T	T	T
F	T	T	T	F	F	T	T	T
F	T	F	T	F	F	T	T	T
F	F	T	T	T	F	T	T	T
F	F	F	T	T	F	T	T	T

4.2 b

List the four main representation schemas learnt from this course and give a typical example for each of them.

1 Logic Representation:

Typical examples of logical representation are Predicate Logic and Propositional logic.

2 Procedural Representation:

Production systems are a typical example of procedural representations.A production system has three components:

A database of facts (also called data memories, or working memories)

A set of production rules (also called condition-action rules, or production memory)

An interpreter (also called inference engine)

3 Network Representation:

Semantic networks are a typical example of network representation.

4 Structured Representation:

Frames are one of the most common data types of structured representations

5 Question

There are three men: two good men and one bad man. One of the good men is rich and the other is poor. Each of the two good men only makes statements which are true and the bad man only make statements which are false. Write down a statement which will guarantee that one of these persons who can make such a statement must be the rich good man. Please critically justify your answer by proving that it can be only made by the rich good man, but neither the poor good man, nor the bad man. [10 marks]

Statement:

"i am Rich if and only if i am telling the truth".

suppose:

p = i am rich

q= i am telling the truth

By converting that Statement to propositional logic we will have:

$p \leftrightarrow q$

this is a bi-conditional statement is only TRUE when both p and q have same truth values.

p	q	$p \leftrightarrow q$
T	T	T
T	F	F
F	T	F
F	F	T

we can divide this into 3 cases:

case1: [Good person who is poor and truth teller]

From the above statement we can deduce that "The good person who is poor and tells truth" cannot make such statement because the fact that he is Poor and truth teller , and statement says "i am rich if and only if i am telling the truth". in that case this Statement would be INCONSISTENT. he cant be Rich and truth teller and poor at the same time.

case 2:[Good person who is rich and truth teller]

he always tells the truth so the statement has to be True and there are just 2 scenarios when statement is true.

1: when p=T and q = T
2: when p=F and q=F

As we know that this person is Rich and always tells truth so in this case both Values needs to be true for the statement, so scenario 1: when p=T and q=T is according to our statement and we can conclude that the above statement was made by RICH GOOD person.
the second scenario [when p=f and q=f] cannot be accepted because our above

statement says "if and only if i am telling truth" and in second scenario q=f which means he is telling Lie.

p	q	p↔q
T	T	T
T	F	F
F	T	F
F	F	T

case 3:[Bad person who lies] as we know this person always lies , whatever he says is FALSE so

p↔q=F

there are 2 scenarios when p↔q=F

1: when p=T and q=F

2: when p=F and q=T

we know that he is a liar so the first scenario is most suitable where p=T and q=F which makes above statement false.

6 Question

(a) State the axioms of probability.

1: The probability of any given proposition is no less than 0 and no bigger than 1:

$$0 \leqslant P(A) \leqslant 1$$

2: Logical True propositions have probability of 1, and logical false propositions have probability of 0:

$$P(True) = 1'and'P(False) = 0$$

3: The probability of the disjunction of A and B is given by:

$$P(A \vee B) = P(A) + P(B) - P(A \wedge B)$$

(b) Explain briefly why reasoning using probabilistic data is important in artificial intelligence.

Uncertainty in knowledge is the main cause of using probabilistic data. So to represent uncertain knowledge, where we are not sure about predicates, we need uncertain reasoning or probabilistic reasoning. Probability provides a way of summarizing the uncertainty that comes from our laziness and ignorance. Uncertainty can be defined as the lack of the exact knowledge that would enable us to reach a perfectly reliable conclusion. [4 marks]

(c) A disease D causes two symptoms S1 and S2 in an individual with probabilities P(S1|D), P(S2|D), by two independent mechanisms. Explain why the occurrence of the symptoms should be treated as independent only if it is known that the disease is present i.e. the relation P(S1, S2|D) = P(S1|D).P(S2|D)

holds, but not P(S1, S2) = P(S1).P(S2). [5 marks]

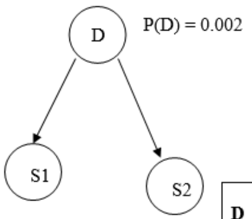
$$P(S1,S2|D) = P(S1|D).P(S2|D)$$

This equation expresses the Conditional Independence of the symptoms S1 and S2, given Disease D. Each symptom is directly caused by Disease D, but neither has direct effect on the other. Conditional independence is most basic and robust form of knowledge about uncertain environments.

$$P(S1,S2) = P(S1).P(S2)$$

This is called Independence. Symptoms S1,S2 are independent of Disease D. The independence assertions can help in reducing the size of the domain representation and the complexity of the inference problem.

(d) Calculate the probability of a person having a disease D given that they are showing both symptoms S1 and S2, from the Bayes Network of Figure 2. [13 marks]



D	P(S2|D)
T	0.7
F	0.04

D	P(S1|D)
T	0.3
F	0.01

Figure 2

$$P(D|S1,S2) = \frac{P(D,S1,S2)}{P(S1,S2)}$$

Solving the Denominator

$$P(S1,S2) = P(D,S1,S2) + P(\sim D,S1,S2)$$

$$P(D,S1,S2) = P(D) \times P(S1|D) \times P(S2|D)$$

$$P(\sim D,S1,S2) = P(\sim D) \times P(S1|\sim D) \times P(S2|\sim D)$$

Inserting values into equations

$$P(D,S1,S2) = 0.002 \times 0.3 \times 0.7 = 4.2 \times 10^{-4} = 0.00042$$

$$P(\sim D,S1,S2) = (1-0.002) \times 0.01 \times 0.04 = 3.992 \times 10^{-4} = 0.0003992$$

original Equation

$$P(D|S1,S2) = \frac{P(D,S1,S2)}{P(S1,S2)}$$

hence

$$P(D|S1,S2) = \frac{P(D,S1,S2)}{P(D,S1,S2) + P(\sim D,S1,S2)}$$

$$P(D|S1,S2) = \frac{P(D,S1,S2)}{P(D) \times P(S1|D) \times P(S2|D) + P(\sim D) \times P(S1|\sim D) \times P(S2|\sim D)}$$

values are inserted into equation

$$P(D|S1,S2) = \frac{0.002 \times 0.3 \times 0.7}{(0.002 \times 0.3 \times 0.7) + ((1 - 0.002) \times 0.01 \times 0.04)} = \frac{4.2 \times 10^{-4}}{8.192 \times 10^{-4}} = \frac{0.00042}{0.0008192}$$

$$P(D|S1,S2) = 5.1269 \times 10^{-1} = 0.51269$$

7 Date

December 5, 2019