

Feedback provider name: Omar Elbaghdadi
Feedback provider student number: 12660256
Feedback recipient name: Jan Zuiderveld

Instructions:

Take some time to read the other group's report or blog (about 30-45 minutes). After that, use this feedback form to note your comments.

For each criterium (row), make clear which of the options apply, by e.g. underlining it or marking it in color. For each section, explain in a couple of sentences why you have given the rating, and what the other group could do to improve on those aspects.

75 minutes after the start of the tutorial session (12:15 or 14:15), get back together in the breakout room to share your feedback. You might both take 5-10 minutes to read the feedback, and then ask each other for additional explanation if something is still unclear. In addition to sharing the feedback with the other person, upload a PDF version of this form in Canvas.

- Presentation. Clear, legible, structured, use design elements (e.g. media) effectively.

Clarity of writing needs to be improved	Clarity ok, but could be further improved	Excellent clarity of writing
Use of design elements (media / headings) needs to be improved	Ok use of design elements, but could be further improved	Excellent use of design elements

explanation / possibility for improvement:

- so far no images (but also reasonable as no results yet). However, in a blog post more images is almost always a good idea.
- In your writing, decide how formal / informal you want to be. For example in some sentences many complex terms are used (e.g. first sentence), but other times the writing is very informal, e.g. "And it doesn't stop there".
- Instead of having a section called Introduction, maybe name it to Q-learning or something like that. That's more informative in my opinion.

- Motivation and research question. What is the central question and why is it important?

Motivation or research questions need to be improved	Motivation & research questions ok, but could be improved further	Excellent motivation and research questions
--	---	---

explanation / possibility for improvement:

- I think the motivation for the importance of this study is not clear enough. Now the applications of Q learning are just highlighted:

"More and more new applications are being explored and built in robotics, autonomous driving, just to name a few. With all that being said, the bottleneck of semi-gradient methods and its variations with experience replay,

target networks remains an important question for this area."
 But this doesn't tell us why studying DQN techniques is important TO Q-learning.

- Explanation of the algorithm and techniques used. Explained intuitively where possible.

Missing formal description (e.g. algorithm, equation, pseudocode)	Formal description ok but could be improved	Excellent formal description
Missing explanation/intuition	Explanation/intuition ok but could be improved	Excellent explanation / intuition.

explanation / possibility for improvement:

- First paragraph of introduction: sentence is quite complicated. If you want to do a blog post style, it is better to have less complex words and less complicated / long sentences in my opinion. (However, I don't know how formal the TA's want us to be, even in blog post style.)
- Q-learning explanation
 - ... is the target value, which correlates with the Q value --> what do you mean by correlates with the Q value? Vague description imo.
 - This correlation makes it difficult to apply full SGD on a Q function parameterized by w --> why? no explanation
 - Semi-gradient methods assumes that the target is independent of w , by minimizing the mean square Q value error, we can arrive at the semi-gradient version of Q learning weight update rule --> Sentence too long. Have a period somewhere in between.
 - Semi-gradient methods have important advantages. They typically enable faster learning --> compared to what?
 - no clear introduction of function approximation vs tabular and how that makes the learning more unstable, maybe talk about deadly triad?
 - The use of deep neural networks signifies narrowing a gap for the reinforcement learning problem --> what gap?
 - Another benefit of DQN is that it maps an input state, s_t , to the set of legal actions, $A(s_t)$, in one single network, which avoids the cost of the network to scale up with respect to the number of actions, as is the case in pairwise Q-Learning algorithms --> This is not something that DQN introduced I think.
 - Maybe it's good to put more emphasis on the correlation between states close to each other in time and that the target moves after every update. So a clearer motivation for the most important innovations by DQN.
 - At the moment section 3.1 is one really long paragraph. Try to break it up into smaller paragraphs.
 - DQN is model-free could maybe better be stated as DQN uses Q-learning. Q-learning is model-free can then be said when introducing Q-learning.
 - Section 3.3
 - when an update increases $Q(s_t, a_t)$ also increases $Q(s_{t+1}, a)$ for all a and, hence, an increase in the target value y --> strange wording. Also, if $Q(s_t, a_t)$ increases, it doesn't mean that $Q(s_{t+1}, a)$ increases.
 - Otherwise quite clear explanations of experience replay and target networks

- Experimental design. How and why were experiments set up?

No comparison at all, or major improvement needed	Comparison to baseline, relevancy of baseline or environment not sufficiently clear	Excellent comparison, clear why this environment and baseline was chosen
single training run on single environment	multiple runs and/or multiple environments	
unclear how hyperparameters are chosen and which are used	clear how hyperparameters are chosen and which are used	

explanation / possibility for improvement:

- Maybe explicitly say what the baseline is (DQN without tricks)
- No clear motivation for environments
- We decided not to tune the discount factor due to a lack of time and limited influence --> Can you motivate limited influence? In our experiments we found discount_factor to be a major influence to results for example. It's fine to not do a hyperparameter search over it but try to substantiate any claims you make.
- Discounting previous (future*) rewards is less influential when correlation is already being disrupted by experience replay -> This is also just a claim, no evidence.
- Nice that you calculate statistical power of your tests. But this section could maybe be shortened. E.g. just mention we use statistical power and find that we need 26 samples.
- Convergence will be determined using thresholding of loss difference --> Maybe give some motivation for this choice? There are other ways to define divergence as well.

- Results and conclusion. Are results clearly presented & conclusions warranted.

Graphs and tables unclear - e.g. no axis labels, or unclear what quantity is shown	Graphs and tables clearly labeled, clear how quantities are calculated
No spread of results shown, or unclear what	Clear indication of reliability of results

is shown	(including clear definition of what is shown)
Unclear experimental procedure (number of runs, which environments & baselines)	Clear how experiments were performed clear how hyperparameters are chosen and which are use
No conclusions, or not sufficiently supported	Conclusions are sufficiently supported. Strength of conclusions take limitation of experiments into account.

explanation / possibility for improvement:

- Results section is not present yet.

- Credit. Clear mention of use of resources (environment, algorithms) by others.

Credit is not given everywhere where due, or own contribution of group is unclear.	Own contribution is clear, and credit is given where due.
--	---

explanation / possibility for improvement:

- Good job to mention where the code is from.

- Group size

For groups of 5: project does not exceed expectations of 4-people groups	Clear additional effort compared to minimal requirements (e.g. more than 2 environments, more than 2 algorithms compared, above average amount of own implementation)
--	---

explanation / possibility for improvement:

- Other feedback (optional)
 - Jan told me that his group wants to do it in blog post format, but until now they only have it written out in latex. I would recommend to use markdown to make it look like a blog post. Github can be used to work together.