QCAT, 1 Technology Court,
Brisbane, Queensland 4069
Australia
T (07) 3327 4008 • **ABN** 41 687 119 230
12 March 2020.

To, Professor Benoît Clement, Professor Karl Sammut

Re: PhD review report for Reinforcement Learning and Transfer of Adaptive Control Parameters

Dear Professor Benoît Clement, Professor Karl Sammut,

Please accept this letter as my review report of the PhD Thesis "REINFORCEMENT LEARNING AND TRANSFER OF ADAPTIVE CONTROL PARAMETERS FOR IMPROVED ROBUSTNESS TO UNOBSERVABLE CURRENT DISTURBANCE" submitted by Mr. Thomas Chaffre towards the fulfilment of his candidature.

Summary:

The thesis presented here identifies the gaps in current research and capability for control of Autonomous Underwater Vehicles (AUVs) in the presence of unmodeled disturbances. By carefully identifying the literature of work both on classical control as well as recently proposed adaptive and machine learning approaches, this thesis crafts a strategy for formulating the problem in terms of Adaptive Pole-Placement Control performed using state of the art approach in Deep Reinforcement Learning (DRL). Integration of a novel experience replay buffering technique is also shown to improve the performance of the results. The thesis supports the claims made by performing analysis and presenting arguments, along with empirical results in terms of performance plots in simulation and on physical hardware. The results are often validated by peer-reviewed publications arising from this thesis, additionally validating the contributions made by independent reviewers.

Overall, the novelty of the approach, the relevance of the application, the depth of discussion, and the experimental results are above the benchmark for a successful PhD thesis, and thus, I have no trouble recommending acceptance of this thesis towards a successful doctoral program with minor edits.

Writing and presentation:

The thesis is written very well. The language is clear; the arguments flow very well, from the explanation of the challenges to the development of a plausible research approach.
The figures are clean and easy to read. The plots and diagrams are relevant to understand the text. However, there could be better consistency in the formatting of the images. E.g., Fig 2.12 and 2.13 are presented as a 3-D bar plot which only adds a little more information than a 2D format which the rest of the thesis uses. However, this is a minor and individual fix.
Additionally, there were quite a few grammatical errors throughout the thesis that can be fixed by simple proofreading. e.g." ... never been explored by neither controllers during training ... " should be "never been explored by either controller... " (p62 line2)
While these are mildly distracting, they are not critical for comprehension, however, should be fixed in the final submission.
Another minor comment in revising the text would be to ensure that the abbreviations are reintroduced sufficiently enough throughout the text, especially when they are newly introduced by the author, e.g., LB learning based, which is sometimes not referred to in the plot. Another example is the usage of ADPP without definition in Fig.3.11 (b) which I assume stands for adaptive pole placement. Chapter 4 suffers the most in this situation. A suggestion would be to provide an index of all the abbreviations used in work for quick reference. Also, for the purposes of the thesis, the subtle distinctions between the usage of AUVs and UUVs in application contexts are not relevant, and it would be better to stick with one.

Literature Review and positioning of the thesis w.r.t prior work:

The thesis presents a good literature review of both the classical and machine learning approaches taken to address the challenge of uncertainty by unknown system (and process) dynamics for AUVs. This spans from adaptive control with pole placement approaches to learning-based approaches with parameter tuning. I was glad to see a good background on the introduction of Reinforcement Learning (RL), from a basic formulation of the bellman equations all the way to current algorithms being used in Deep Reinforcement Learning (DRL).
A good indication of the principled approach taken by the researcher is the presentation of the experimental evaluation of two different approaches in chapter 3, where the comparison of naive approaches both from RL and a simple PID controller clearly lays bare the necessity of developing modelled learning approaches towards the solution.

Experimental comparison methods:

The approach presented adaptive pole placement using RL in this thesis clearly outperforms both the naive RL (with training on direct control outputs) as well as PID feedback.
There is a discussion about how the proposed approach lines up with Model Predictive Control (MPC) in principle, but it would have been good to see an experimental or theoretical derivation of the claim, even if for a simple scenario.

In the experimental hardware scenario, it was surprising to find that the learning-based approach performed better than the model-based control without disturbance. As learning-based approaches tend to model the system dynamics in a latent space to make inferences, there are often misrepresentations inherent in the learned model, and the performance is lower than a well-designed model, especially if the system is fully observable and controllable. If there is a disturbance, then it is logical that the models are not representative, and learning-based approaches might be more robust. The result shows that the absence of a disturbance for a fully observable and controllable ROV without wind conditions, therefore, requires more explanation and probing to understand the underlying insight. Is this an artefact of measurement, or is something more interesting happening underneath?

On the whole, it was good to see that the results in hardware supported the argument that learning-based approaches are necessary to improve station keeping and tracking in the presence of disturbances.

Overall conclusions and observations:

The thesis exposes the challenges of controlling AUVs under unknown process disturbances. By proposing and executing a novel approach of integrating Deep-RL algorithms to perform adaptive pole placement, challenges, and limitations both from the modelling perspective (controllability assumption and observation assumption) and learning perspective (sensitivity to suitable representation and choice of hyperparameter) were unearthed. The thesis provides concrete steps towards improving these challenges and perspectives that are undoubtedly useful for future researchers to build upon and develop the next generation of algorithms and systems to solve a critical real-world challenge.

In that sense, the thesis as presented has made a good contribution to pushing the state of the art of understanding in research.

Warm Regards,

Dr. Tirthankar Bandyopadhyay
Senior Research Scientist,
Robotics and Autonomous Systems Group
DATA61, CSIRO
tirtha.bandy@csiro.au
(07) 3327 4008