Point-by-Point Response to Reviewers

Date: 03-Jan-2025

Manuscript: **An Empirical Study of the Naive REINFORCE Algorithm for Predictive Maintenance**

Dear Editor-in-Chief,
Discover Applied Sciences

Thank you for providing us an opportunity to revise and submit our manuscript, “An Empirical Study of the Naïve REINFORCE Algorithm for Predictive Maintenance”.

We have addressed all the comments and it has greatly helped enhance our manuscript. We have formally thanked and acknowledged them in our manuscript.

In this point-by-point response document we address all the observations and provide page and section references so that the revisions can be traced back to the revised manuscript. In the revised manuscript the changes are highlighted in light-blue.

Summary:

- Assistant Editor: Article Highlights were missing. Added to manuscript.
- Reviewer # 1: 6 comments (9 total observations), addressed on: Pages 2 to 5.
- Reviewer # 2: 9 comments (12 total observations), addressed on: Pages 6 to 10.

Kind regards,

Dr. Satish Kumar

Symbiosis International (Deemed University)

Reviewer #1 Comments

We thank the Reviewer for the positive and encouraging comments, along with the meticulous review of our manuscript. The Reviewer suggested adding hyperparameter analysis, training time reduction strategies and comparison to ML – all of these are pertinent when conducting research in the field of reinforcement learning (RL). We now cover these and other technical suggestions, and this has helped improved the technical content of our manuscript significantly.

Comment 1	While the study's focus on "untuned" RL models is its core premise, exploring minor variations in key hyperparameters could provide richer insights. For instance, a broader evaluation of learning rates for REINFORCE may clarify whether its exceptional performance stems from unique dynamics or optimal settings.
Response 1	This is a significant suggestion and has helped enrich the manuscript technically. Your suggestion on analyzing the impact on variations of the hyperparameters on REINFORCE added to the novelty of our research and we devoted an entire new sub-section for this. We first studied existing research to identify which hyperparameters to analyze. We analyzed learning-rate, the discount factor γ and network activation functions (Tanh and the ReLU). We then designed and ran experiments to analyzed the results. Interesting insights were generated and discussed the findings in the Discussion Section. Several new research articles were studied and cited in our manuscript. Addressed on Pages 29, 30 and 31 in section "Sensitivity analysis of hyper parameters" and the Discussion Section on page 31, 32 and 33. We added Figures 23, 24, 25 and 26.

Comment 2	REINFORCE's training time is reported as significantly longer than the other algorithms, with considerable variance. However, there is no discussion of strategies to reduce its training time. I suggest adding a cost-benefit analysis comparing training time to achieve precision or recall.

Response 2

Your previous suggestion on analyzing hyperparameters, helped us unveil some of these strategies. This also helped us review some established research on this topic, specifically François-Lavet et al. (2016) and Eimer et al. (2023).

We mention strategies around the discounting rate, the learning rate and the use of the ReLU activation layer.

Addressed on Pages 30, 31 and 33. Section "Sensitivity analysis of hyper parameters", "Impact of hyperparameter setting on training time" and "Compute cost of performance improvement". Fig. 25 and Fig. 26.

Eimer, T., Lindauer, M., Raileanu, R.: *Hyperparameters In Reinforcement Learning And How To Tune Them*. In: International Conference on Machine Learning, (2023). PMLR

Francois-Lavet, V., Fonteneau, R., Ernst, D.: *How To Discount Deep Reinforcement Learning: Towards New Dynamic Strategies*. CoRR abs/1512.02011 (2016)

Comment 3 The REINFORCE algorithm reported pool recall, especially in complex multivariate states' experiments. The recall is critical in industrial maintenance tasks because failing to replace a tool at the right time can lead to catastrophic failure or significant downtime. Therefore, this poses a risk to reliability in high-stakes environments.

Response 3

We agree. Our focus was on precision which is driven by lower false-positives (FPs) i.e. reducing unnecessary replacements. However, we agree, in general industrial scenarios, the cost of tool is much "lower" compared to cost-of-quality.

We have therefore modified our text and state the importance of considering the metric on the basis of the industrial application and considering F1 as a more balanced measure. We provide suggestions on using suitable value for F-beta ($\beta = 2.0$) when importance to recall is desired. The abstract is also modified to include recall metrics.

We modified the text to state that "for this research we consider $\beta = 0.5$". This is probably applicable in situations where the actual process of replacing part itself is expensive and a considered a hazard for the maintenance staff, for example maybe wind-turbine maintenance.

Addressed on Page 20: Evaluation metrics.

Comment 4 While the concept of testing untuned models makes the results more accessible to practitioners, it also limits the study's depth. Practically tuned models are preferred for deployment. Also, comparing untuned versions ignores the fact that other algorithms (e.g., PPO or DQN) might outperform REINFORCE when tuned.

Response 4 This is correct and is a point we had missed stating. We now mention that advanced algorithms provide a richer set of hyperparameters and that the tuned versions of these algorithms will perform better than the naive REINFORCE. However, we admit that a deeper study by tuning these algorithms is a limitation of our scope and is an important subject for future research.

Addressed on Page 34, Discussion Section under "Limitations and future scope for research".

Comment 5 The paper attributes REINFORCE's surprising performance to factors like activation functions (ReLU vs. Tanh) and learning rate. Still, there is no discussion of architectural simplicity as a driver for better performance in smaller data environments. I suggest including baseline comparisons against simpler supervised learning models to contextualize the value of RL.

Response 5 Thank you for highlighting this. We analyze the view of established researchers on architectural aspects of RL implementations as well as REINFORCE's architectural simplicity as a possible reason for its better performance.

We surveyed **12 articles** that have used supervised machine learning (ML) to solve a similar predictive maintenance (PdM) problem. A new table was added (Table 1), where we list the ML technique used, the type of time-series problem (univariate or multi-feature) it was addressing and the actual PdM use-case it solves. In the Discussion section, we briefly compare the RL and ML methods.

Addressed on Page 7, Table 1 and Page 33, Discussion Section.

Comment 6 Although statistical metrics and tests are provided, the paper does not offer confidence intervals for most metrics, limiting interpretation of results' robustness. Also, variability in performance across datasets is not well analyzed or explained.

Response 6 We have now added 95% confidence intervals to all the results tables in the main text. The new plots from the Sensitivity Analysis contain 95% confidence intervals. All the performance plots already had error bars.

Revised tables in the Results Section: Table 6 pg. 21, Table 7 pg. 23, Table 8 pg. 24, Table 9 pg. 25 and Table 10, pg. 27. Sensitivity Analysis section Figures 23 and 24 on Page 30.

The variability in performance across datasets is a result of the high variation in features. We have now explanation on in the 'Actual tool wear data' sub-section of the 'Implementation details' section and added plots to show this variation, see Fig. 6, Page 12.

---------- End of "Reviewer #1 Point-by-point" section ----------

Reviewer #2 Comments

We would like to thank the Reviewer for helpful suggestions and highlighting multiple gaps in our research. As research addressed for industrial practitioners, we had indeed missed certain aspects, like cost effectiveness. Other areas of improvement were suggested – for example we had completely missed mentioning limitations of our research scope. We have addressed each of these and this has enhanced the manuscript, thus adding value to the RL research community.

Comment 1	Please include the stats of the paper found and also the methodology employed in this paper.
Response 1	We have added the statistics in the Literature Review section and added a figure depicting the volume of articles over years. In the text, we have explained the techniques used in the papers – for both newly added papers as well the previous ones. For example, how noise in RL is studied in empirical research in Eimer et al. (2023) and how hyperparameters were selected in similar empirical research articles, by Shala et al. (2022) and François-Lavet et al. (2016). 11 new research articles were added with the methods they use in Table 1. Addressed on Pages 4 and 5 and Table 1 on Page 7. Eimer, T., Lindauer, M., Raileanu, R.: *Hyperparameters In Reinforcement Learning And How To Tune Them*. In: International Conference on Machine Learning, (2023). PMLR Shala, G., Arango, S.P., Biedenkapp, A., Hutter, F.: AutoRL-Bench 1.0. In: *Sixth Workshop on Meta-Learning at the Conference on Neural Information Processing Systems* (2022). Francois-Lavet, V., Fonteneau, R., Ernst, D.: *How To Discount Deep Reinforcement Learning: Towards New Dynamic Strategies*. CoRR abs/1512.02011 (2016)

Comment 2	There is no discussion on the cost effectiveness of the method. What is the computational complexity? What is the runtime? Please include such discussions.
Response 2	Since this paper was addressed toward industrial practitioners, these are important discussions items. Thank you for highlighting this. As suggested, we have now covered these topics in the Discussion Section. **Computational complexity**: We discuss design of neural network architectures for RL and REINFORCE (Page 33). On Page 30, we analyze the impact of hyperparameter setting on training time and analyze the cost of compute, which is a function of training time, on performance improvement. On Page 31, a new figure is added to visualize this. We also discuss the strategies might help reduce training time and therefore cost of training on Page 33. **Runtime**: Runtime is a function of the model size. We conducted **additional experiments** and added the average size of models produced. These are covered in Section 4.6.1 **Training times and model byte size** (Page 28). On Page 33, we discuss the significance of the small model sizes – since the model is computationally light and model sizes are about 120 KB, they are suitable used on **embedded IoT** devices and housed near the machines. This is a significant benefit for creating **Industry 4.0** predictive maintenance solutions, producing inferences that are near real time. Addressed on: Pages 28, 30, 31 and 33. Figure 26
Comment 3	To have an unbiased view in the paper, there should be some discussions on the limitations of the method.
Response 3	We agree and a separate standout section has now been added to the Discussion Section, Page 34 titled “Limitations and Future Scope for Research”. We identified 3 limitations and 3 opportunities for future research. Additionally, we have now cover the drawbacks of the REINFORCE algorithm in the Discussion Section such as premature convergence, high variance, the need for large sample sizes and extended training times. Addressed on: Pages 32 and 33 and Page 34

Comment 4	Neither the novelty nor the uniqueness of the research is established.
Response 4	This comment helped us take serious note of our writing to bring clarity and surface the contributions. We **completely re-wrote** the Literature Research section to understand existing research and identify and highlight **research gaps**. Our work is directed toward industrial predictive maintenance (PdM). It touches the fields of AutoRL and therefore HPO (hyperparameter optimization) and is essentially an empirical study of the simple REINFORCE algorithm that is often ignored for real world applications. Our first contribution is related to the study of untuned RL algorithms for a PdM problem. This is a first, small step contribution toward AutoRL for PdM. The multiple surveys we studied on AutoRL indicate unanimously that there is no one solution and therefore a **domain-specific study** is important. Most existing RL experiments are always on OpenAI Gym environments. It has been shown by established researchers that implementing robust RL algorithms is complex, due to the closed-loop nature of RL. Therefore, our study on **hyperparameter sensitivity and interaction analysis** – for the **REINFORCE** algorithm is the second contribution. Thirdly, none of the AutoRL or HPO empirical studies touched network architecture elements. Our small contribution has been the study of the main hyperparameter, the activation function. The StableBaselines open-source implementations use Tanh, while we showed that ReLU is an efficient activation function, both for average episodic rewards as well as time to train.
Comment 5	Authors need to add more latest references from the years 2022 and 2024.
Response 5	Our original manuscript had 30 references. Our work to address the Reviewer comments, resulted in addition of **21 new** references. Of the 51 references now – **16** cover the 2022-2024 period: 2024: 4, 2023: 5, 2022: 7 Addressed in: The Reference section – Page 43-47.

Comment 6 Abstract needs to relook and highlight the scope and then add what is the aim/Objective of the paper, also highlight the numerical Findings and compared to existing works to justify that the training set model works better and what is the overall analysis.

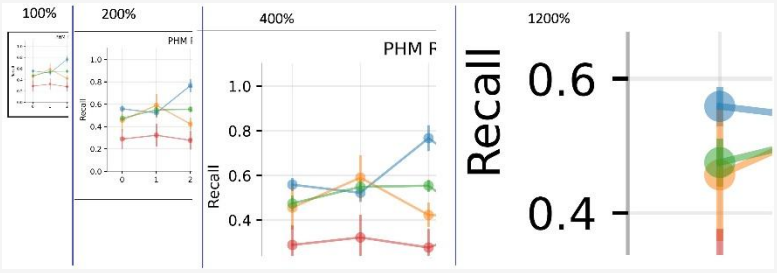
Response 6 We have now revised the Abstract. Important changes include hyperparameter sensitivity and interaction analysis that we conducted to better understand the REINFORCE dynamics. Since our research is aimed at assisting the AutoRL field – we mention that this research will encourage new design approaches for the automatic identification of optimum algorithm-hyperparameter combinations.

Comment 7 Suggested to relook the conclusion section and highlight the open issue for further research contribution. The quality of the figures and tables need to be checked.

Response 7 The Conclusion section was significantly re-written to address this. We carry the findings we discuss in the Discussion Section from the additional experiments we performed. These covered the three hyperparameters we studied learning rate and the discount factor γ and the activation layer functions.

Open research questions and opportunities for future research are carried into the Conclusion section from the section “Limitations and future scope for research”.

We have recreated many images in vector format (PDF) and ensure extremely high quality ensuring clarity beyond 400% zooms.



Addressed on Pages 34 and 35. Quality of figures throughout the paper.

Comment 8	The technical contribution of this research is not adequately described in the abstract. I advise rewriting it.
Response 8	Thank you for highlighting this – we have rewritten the Abstract to include the hyperparameter sensitivity and interaction analysis and impact to the AutoRL field.
Comment 9	The methods part is poorly designed and needs improvement to include more evidence on the adequacy of the research procedure.
Response 9	As a paper on empirical study – this was important to being a scientific structure to the study. We modified two main areas of the manuscript. First, we have **re-written** the initial foundation of the Methodology section. We start the section with an explanation of the research approach we employed, namely "**Exploratory Research**". We linked and mapped our phases to the original **empirical cycle** proposed by **Adriaan D De Groot**. We added a **new Table 2**, that maps our activities to the original empirical cycle. Second, we ensure we cite references to researchers who have used a similar methodology for example, we mention that the study of effects of noise on RL was as suggested by Eimer et al. (2023) on Page 12 and on Page 29 for selection of the hyperparameters we chose to study. Addressed on Pages 9, 10 and 12, 29. Table 2, Page 10. Eimer, T., Lindauer, M., Raileanu, R.: *Hyperparameters In Reinforcement Learning And How To Tune Them*. In: International Conference on Machine Learning, (2023). PMLR

---------- End of "Reviewer #2 Point-by-point" section ----------