# Dynamic Sparsity for Robust Preference-Based Reinforcement Learning

Calarina Muslimani[1], Bram Grooten[2], Deepak R.S. Mamillapalli[1], Mykola Pechenizkiy[2], Decebal C. Mocanu[3], and Matthew E. Taylor[1]

[1] University of Alberta
[2] Eindhoven University of Technology
[3] University of Luxembourg

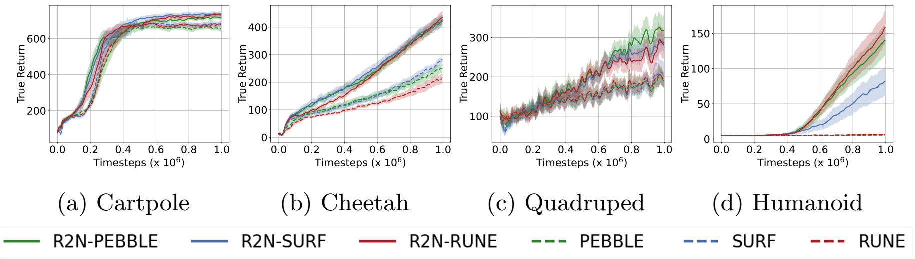
**Keywords:** Deep Reinforcement Learning · RLHF · Sparse Neural Networks

## 1 Introduction

For autonomous robots to integrate into human-centered environments, agents should learn to adapt to human preferences. As the world is full of distracting information, we aim to build reinforcement learning (RL) agents that *learn a reward function* from human feedback, while being robust against irrelevant noise. This work proposes R2N (*Robust-to-Noise*), the first preference-based RL (PbRL) algorithm that leverages dynamic sparse training to learn robust reward models. We demonstrate that R2N can adapt the sparse connectivity of its neural networks to focus on task-relevant features, enabling R2N to outperform state-of-the-art PbRL algorithms in multiple locomotion and control environments.

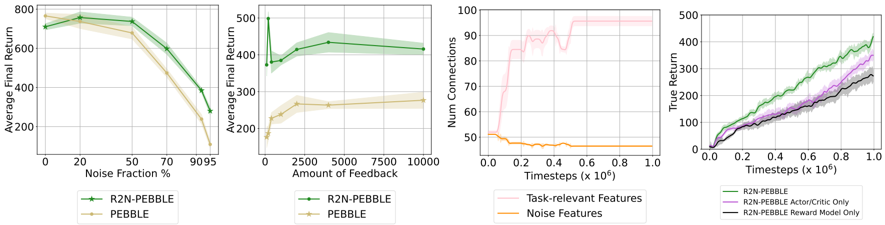
## 2 Method

R2N consists of two main steps. First, at initialization, we randomly prune the input layer of the reward model to a sparsity level $s$. Second, after every $\Delta T$ weight updates, R2N updates the sparse connectivity: a certain fraction $d \in (0, 1)$ of the active weights is dropped (weights with the lowest magnitude), and the same number of inactive weights is activated in new locations. We use RigL [2]



(a) Cartpole (b) Cheetah (c) Quadruped (d) Humanoid

Fig. 1: Learning curves of R2N (solid) and the PbRL baselines (dotted).



(a) Noise Study (b) Feedback Study (c) NN Connections (d) DST Ablation

Fig. 2: Further studies on (a) effects of noise fraction, (b) effects of feedback budget, (c) average number of neural network connections to task-relevant versus noise features in a reward model with R2N, (d) DST component ablation.

to select which inactive connections to grow. We also apply this dynamic sparse training (DST) [7] procedure to the input layers of the actor and critic networks in the RL agent, as done in ANF [3], a noise filtering algorithm for regular RL.

## 3 Experiments

We evaluate R2N on Extremely Noisy Environments [3], an adaptation of the DeepMind Control (DMC) Suite [10] where random irrelevant (noise) features are added to the state space of each environment. We use the tasks Cartpole-swingup and Cheetah-run (with 90% noise added), Quadruped-walk and Humanoid-stand (with 70% noise). To showcase the utility of R2N, we integrate it with three PbRL baselines: PEBBLE [5], SURF [9], and RUNE [6]. We set our hyperparameters to $s = 80\%$, $\Delta T = 100$, and $d = 0.2$. We use a simulated teacher that provides preferences between two trajectory segments according to the true reward function. Although our future work will involve human teachers, simulated feedback has commonly been used in prior works [1,4,5,6,8,9] to avoid the expense of human subject studies.

## 4 Results and Analysis

In Figure 1, we find that adding R2N significantly improved both the sample efficiency and final return of each baseline PbRL algorithm in all environments tested. We perform sensitivity analysis on the Cheetah-run environment. In Figure 2a, we find that for higher noise fractions, R2N maintains a significant improvement in final return. In Figure 2b, we show that R2N outperforms the baseline for all tested feedback budgets. We analyze in Figure 2c the number of connections to each type of input feature; R2N quickly learns to focus its connectivity on task-relevant features. And lastly, in R2N we apply DST to both the reward model and actor/critic networks, so we ablate these components in Figure 2d. Full R2N outperforms variants that apply DST to only one of these.

## References

1. Christiano, P.F., Leike, J., Brown, T., Martic, M., Legg, S., Amodei, D.: Deep Reinforcement Learning from Human Preferences. In: The 31st Conference on Neural Information Processing Systems (2017), URL: https://arxiv.org/abs/1706.03741
2. Evci, U., Gale, T., Menick, J., Castro, P.S., Elsen, E.: Rigging the Lottery: Making All Tickets Winners. In: The 37th International Conference on Machine Learning (2020), URL: https://arxiv.org/abs/1911.11134
3. Grooten, B., Sokar, G., Dohare, S., Mocanu, E., Taylor, M.E., Pechenizkiy, M., Mocanu, D.C.: Automatic Noise Filtering with Dynamic Sparse Training in Deep Reinforcement Learning. In: The 22nd International Conference on Autonomous Agents and Multiagent Systems (AAMAS) (2023), URL: https://arxiv.org/abs/2302.06548
4. Lee, K., Smith, L., Dragan, A., Abbeel, P.: B-Pref: Benchmarking Preference-Based Reinforcement Learning. In: Conference on Neural Information Processing Systems: Track on Datasets and Benchmarks. (2021), URL: https://arxiv.org/abs/2111.03026
5. Lee, K., Smith, L.M., Abbeel, P.: PEBBLE: Feedback-Efficient Interactive Reinforcement Learning via Relabeling Experience and Unsupervised Pre-training. In: The 38th International Conference on Machine Learning (2021), URL: https://arxiv.org/abs/2106.05091
6. Liang, X., Shu, K., Lee, K., Abbeel, P.: Reward Uncertainty for Exploration in Preference-based Reinforcement Learning. In: The 10th International Conference on Learning Representations (2022), URL: https://arxiv.org/abs/2205.12401
7. Mocanu, D.C., Mocanu, E., Stone, P., Nguyen, P.H., Gibescu, M., Liotta, A.: Scalable Training of Artificial Neural Networks with Adaptive Sparse Connectivity inspired by Network Science. Nature communications (2018), URL: https://arxiv.org/abs/1707.04780
8. Muslimani, C., Taylor, M.E.: Leveraging Sub-Optimal Data for Human-in-the-Loop Reinforcement Learning (extended abstract). In: The 23nd International Conference on Autonomous Agents and Multiagent Systems (AAMAS) (2024), URL: https://arxiv.org/abs/2405.00746
9. Park, J., Seo, Y., Shin, J., Lee, H., Abbeel, P., Lee, K.: SURF: Semi-supervised Reward Learning with Data Augmentation for Feedback-efficient Preference-based Reinforcement Learning. In: The 10th International Conference on Learning Representations (2022), URL: https://arxiv.org/abs/2203.10050
10. Tassa, Y., Doron, Y., Muldal, A., Erez, T., Li, Y., Casas, D.d.L., Budden, D., Abdolmaleki, A., Merel, J., Lefrancq, A., et al.: Deepmind Control Suite (2018), URL: https://arxiv.org/abs/1801.00690