Reproducing General Value Function Networks

Subhojeet Pramanik
Department of Computer Science
University of Alberta
Edmonton, Canada
spramanik@ualberta.ca

Abstract

In Reinforcement Learning an agent's state has a huge role in its decision-making processes. Recurrent Neural Networks (RNNs) are commonly used to approximate an agent's state representation in partially observable environments. However, algorithms commonly used to train RNNs such as BackPropagation Through Time (BPTT) require computationally expensive multi-step gradient updates. Further, it is very difficult to incorporate relevant prior knowledge of the agent's objective into the RNN, which might help improve trainability. Recently, General Value Function Networks (GVFN) was proposed as an approach to improve RNN trainability. GVFN restricts each internal state component of an RNN to be a value function prediction about the future. They demonstrated that GVFNs are more robust than traditional RNNs and are capable of learning long temporal dependencies, sometimes with just single-step gradient updates. This paper attempts to reproduce the results present in the original paper. We reproduced the Compass World experiments where an agent learns to predict the probability of facing a wall of a given color in a grid-world environment. Our results confirm that GVFNs in fact show faster convergence and achieve asymptotically better performance than a traditional RNN across various gradient truncation values. Further, GVFNs were shown to be more robust than RNNs, as they show lower standard error throughout the training period. Our results reassure that GVFNs is in fact a better alternative to RNNs in environments with long-temporal dependencies and where domain expertise can be useful.

1 Introduction

RNNs are commonly used to represent an Agent's state in partially observable environments. However, Backpropation Through Time (BPPT), a common algorithm used to train RNNs is prohibitively expensive and requires slow multi-step gradient updates (requiring large truncation values). General Value Function Networks (GVFN) propose an alternative to learn the Agent state by restricting the Agent's state to be a value function prediction about the future. Schlegel et al. [2021] recently demonstrated that GVFNs are highly robust to the choice of the truncation value in multi-step gradient updates in BPPT. Further, they demonstrated that GVFNs are capable of learning long-temporal dependencies in RL prediction problems much better than existing recurrent architectures such as RNN, LSTM and GRU.

This report presents the results of replicating the General Value Function Networks (GVFN) algorithm proposed in Schlegel et al. [2021]. Specifically, we replicate the Compass World experiment presented in Section 10, which show that GVFNs are capable of learning temporal dependencies in RL prediction problems. In the Compass World prediction task, an agent is trained to predict five hard-to-learn GVFs corresponding to the color of the wall an agent is facing in an 8×8 grid world. The environment is partially observable as the agent only observes the color of the grid cell in front of it. We reproduce two algorithms, (1) GVFN-TD and (2) truncated-RNN on the Compass World Prediction tasks.

We implemented all the algorithms using the JAX auto differentiation library [JAX, 2022]. Our project is open-sourced and is available at: https://github.com/subho406/General-Value-Function-Networks-Python. Our results demonstrate that the GVFN-TD algorithm achieves higher asymptotic performance compared to truncated-RNN at various truncation levels. Further, we also found that GVFNs generalize much faster compared to truncated-RNNs even at lower truncation levels. GVFNs were also found to be more robust as they had lower standard error throughout the training period. We reassure that GVFNs are infact a better alternative compared to RNNs and further research in this direction is promising.

2 Results

We trained (1) GVFN-TD and (2) truncated-RNN across various truncation values in the Compass World prediction task. The GVFN-TD algorithm is a GVFN trained using semi-gradient recurrent TD algorithm presented in Schlegel et al. [2021]. Further, the truncated RNN is a vanilla action RNN same as the one used in Schlegel et al. [2021]. All experiments were reproduced using SGD as the optimizer. We used a constant learning rate of 0.01 for training. All other hyperparameters were kept the same as the original paper [Schlegel et al., 2021]. All curves use 30 independent runs and report $\pm$ one standard error.

Figure 1 shows the Root Mean Sqaured Value Error (RMSVE) averaged across the last 200k training steps across various truncation values. For truncation=1, truncated-RNN achieves the lowest asymptotic performance (lower RMSVE is better). Across all truncation values, GVFN-TD achieves higher asymptotic performance compared to the RNN. Further, the performance of GVFN-TD is consistent across all truncation values. Even with a truncation=8, truncated-RNN could not achieve similar results. Figure 2 and 3 presents the learning curves for accuracy and RMSVE across various truncation values for the GVFN-TD and truncated-RNN algorithm. As shown GVFNs can learn noticiably faster compared to RNNs. Even with a single step gradient update, GVFN converges to the optimal solution at around 600k steps. RNNs never reaches this performance even with a truncation value of 8. Finally, GVFN variants are more stable as they have lower standard error across all the steps.

Our results look very similar to [Schlegel et al., 2021] and thus reassure that GVFNs are infact better alternatives compared to RNNs in RL prediction tasks with long-term dependencies. Also, supporting the hypothesis that restricting an agent state to be predictions about the future results in better generalization. We summarize the following advantages of GVFNs over RNNs:

- Higher asymptotic performance and faster convergence.
- Less sensitive to the truncation steps in Back Propagation through time.
- More robust to the choice of initial seed.

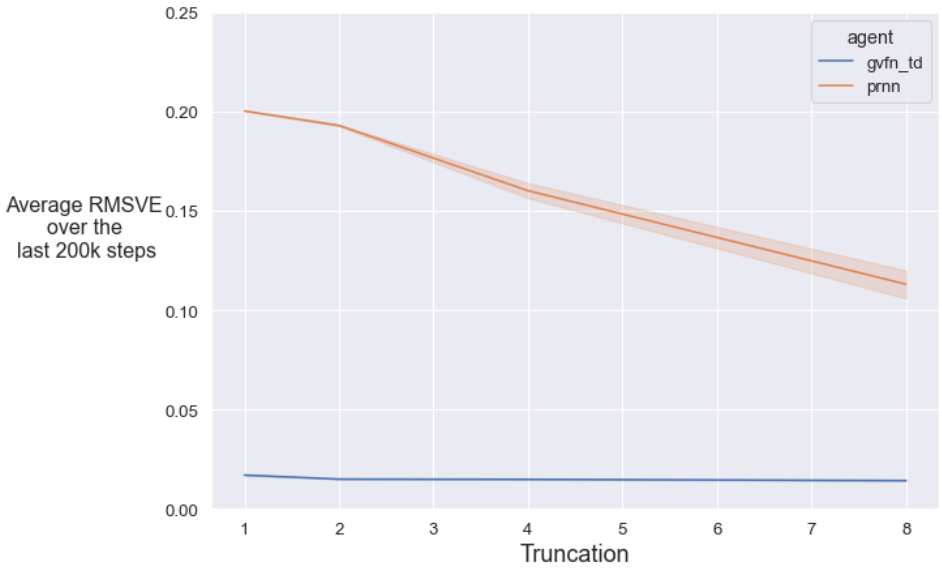

Figure 1: Root Mean Squared Value Error (RMSVE) averaged across the last 200k steps for various truncation values in the Compass World prediction task. Results averaged over 30 runs $\pm$ one standard error. (gvfn_td: GVFN trained with semi-gradient recurrent TD, prnn: truncated action RNN).

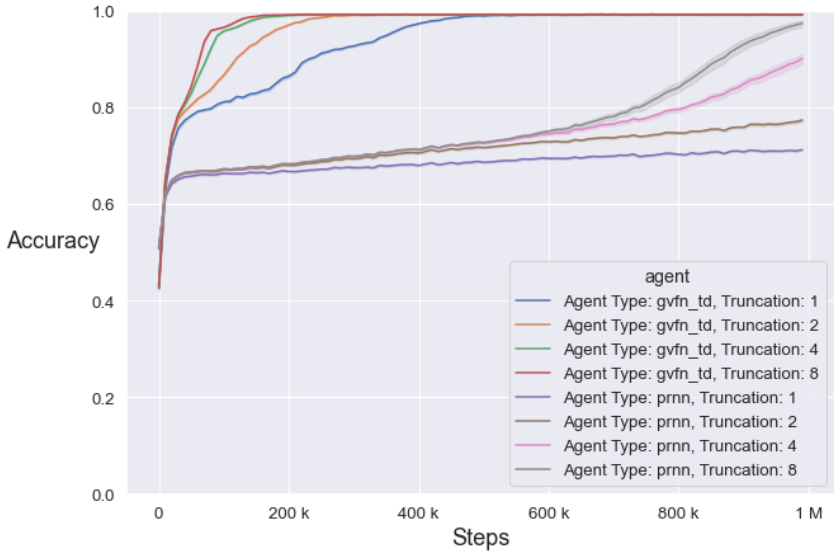

Figure 2: Learning curves for Accuracy in the Compass World Prediction task. We check if the prediction is correct by predicting the color of five with the highest GVF output, where the GVF prediction corresponds to a probability of facing that wall. Results averaged over 30 runs $\pm$ one standard error. (gvfn_td: GVFN trained with semi-gradient recurrent TD, prnn: truncated action RNN)

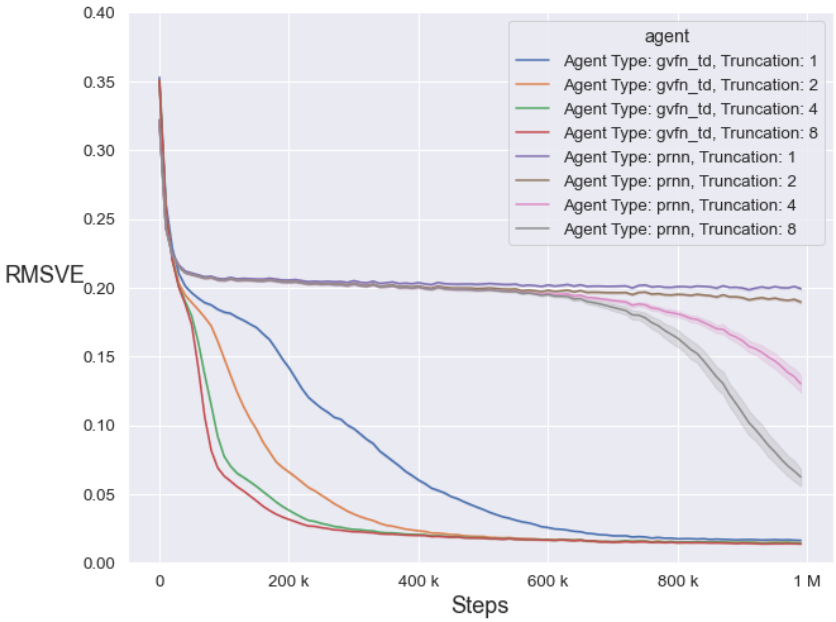

Figure 3: Learning curves for RMSVE in the Compass World Prediction task. Results averaged over 30 runs $\pm$ one standard error. (gvfn_td: GVFN trained with semi-gradient recurrent TD, prnn: truncated action RNN)

Acknowledgements

Huge thanks to Prof. Adam White, and PhD students Matthew Schlegel and Andrew Patterson for their help throughout this project. They were prompt at answering any questions we had about the original paper. Further, we would also like to thank the Computing Science Department at the University of Alberta for providing the resources required to complete the experiments.

References

JAX. Jax: Autograd and xla. 2022. URL `https://jax.readthedocs.io/en/latest/index.html`.

Matthew Schlegel, Andrew Jacobsen, Zaheer Abbas, Andrew Patterson, Adam White, and Martha White. General value function networks. *Journal of Artificial Intelligence Research*, 70:497–543, 2021.