Understanding the effects of discrete representations in Model-Based Reinforcement Learning

An analysis on the effects of categorical latent space world models on the MinAtar Environment

Responsible Professor: Frans A. Oliehoek **Author:** Mihai Mitrea (M.Mitrea@student.tudelft.nl) **Supervisor:** Jinke He

1. Introduction

- Model-Based RL (MBRL) aims to understand and predict the environment's behavior.
- Methods which employ discrete world models have achieved great results [2, 3].
- There is no understanding of why this works, but Hafner et al. [2] provide four hypotheses.

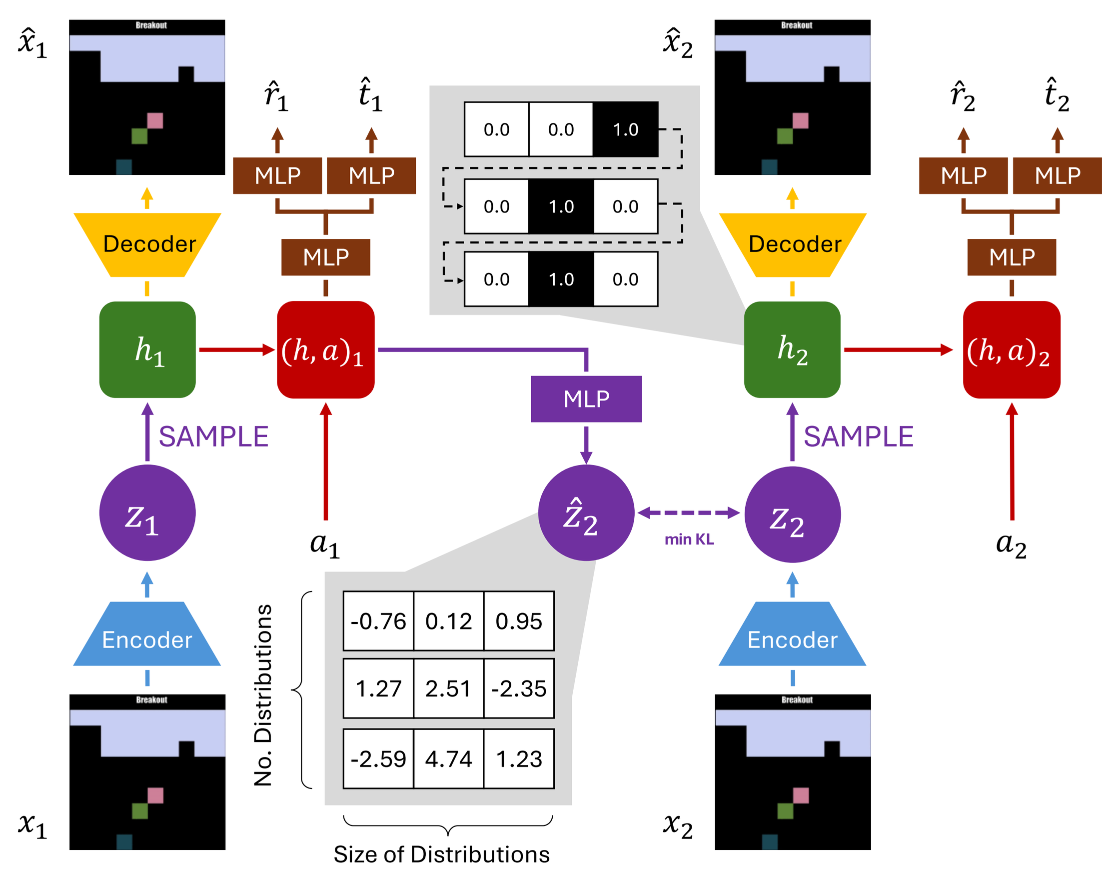
2. Research Questions

How does the discretization of the latent space through categorical distributions affect the world model's performance?

- How does a world model using discrete embeddings compare to a continuous one?
- How does the number and size of the categorical distributions affect the discrete model's performance?
- How does the discretization of the latent space affect the model's ability to generalize?

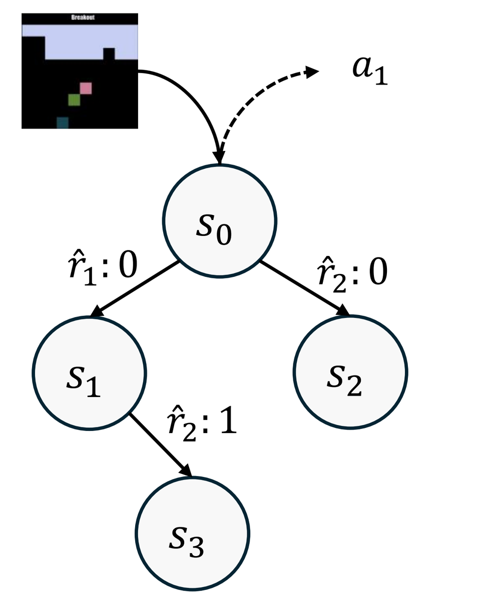
3. World Models

- Architecture of PlaNet [1] and Dreamer [2, 3] but simplified for a deterministic setting.
- Continuous and Discrete latent embeddings are allowed.



4. Experimental Setup

- A simplified Breakout game from the MinAtar environment.
- The models are trained offline, based on an existing experience dataset from a DQN agent.
- An MCTS uses the learned model to plan while interacting with the real environment.
- A model's performance is the mean of multiple episode returns.
- The reported score is the mean and SE of means over three independent models with different seeds.



MCTS planning with the learned model

Step 1: Encode game state

Step 2: Plan using model

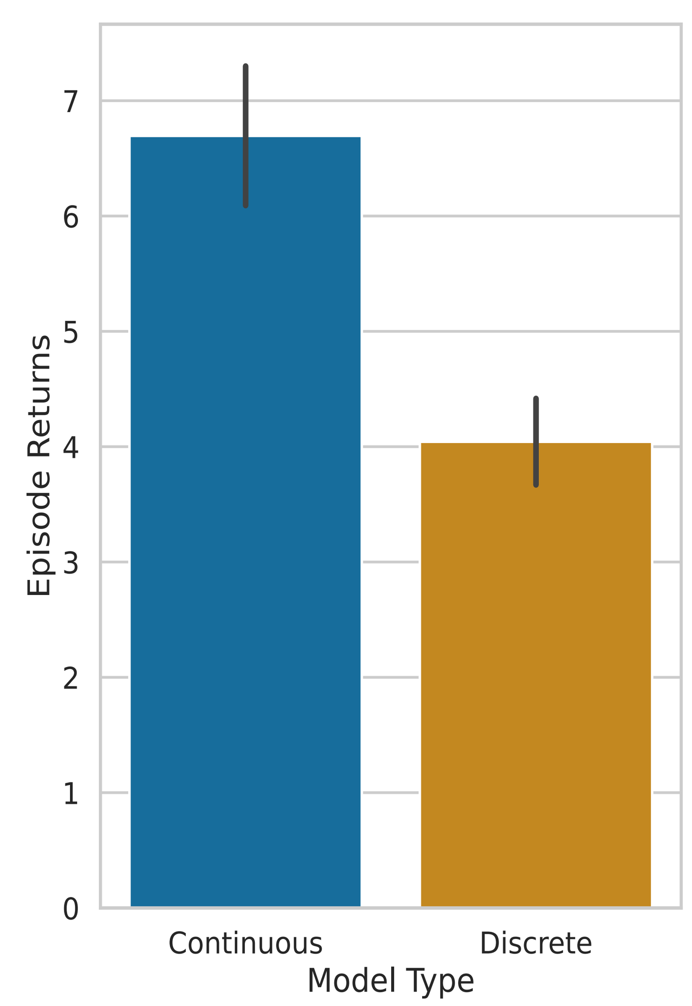
Step 3: Act chosen action

Step 4: Repeat

5. Performance Comparison

The discrete world model is outperformed by a simpler continuous embedding model while being harder to train.

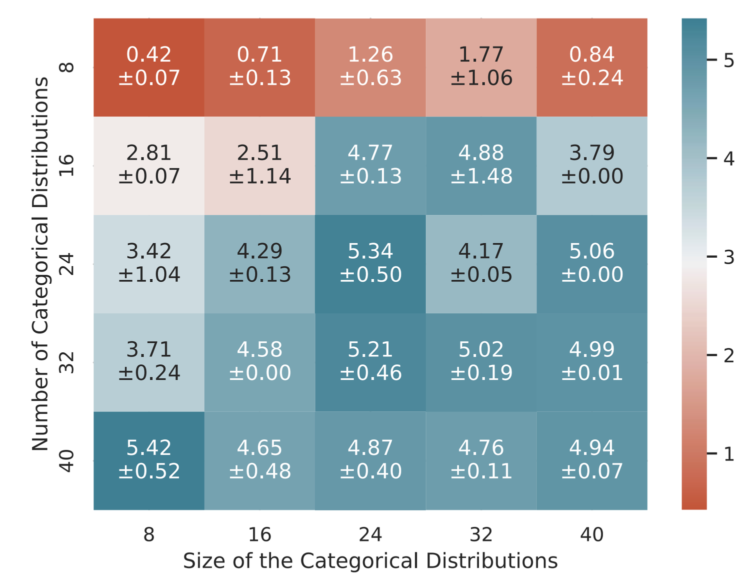
- Both models were fine-tuned in terms of loss function weights and latent space size.
- The discrete model showed a high sensitivity to slight variations in the hyper parameters of the loss function.
- One hypothesis cold be the mismatch between the stochastic nature of the discrete model and the deterministic setting.



6. Ablation Study

Embedding Dimensions

- The number of categorical distributions is the main driver behind the model's performance.
- Past a certain latent size, no improvements are seen.



Generalization Ability

- The discrete world model generalizes better but it sacrifices small gain in more important domains.
- Correctly predicting future states and rewards is more important than continuity.

7. Conclusion

This study finds that in the considered offline deterministic setting:

- The discretization of the latent space does not help the world model achieve better performance.
- The number of categorical distributions is the main driver behind performance.
- The discretization of the latent space improves overall generalization abilities, but at the cost of worse predictions in important metrics.

These **results could show** that two of the four hypothesis proposed by Hafner et al. [2] about DreamerV2's improvements are true:

- The flexibility offered by categorical distributions over Gaussian ones in stochastic environments helps predictions.
- The ability of categorical distributions to more easily represent the "non-smooth aspects of Atari" improves predictions.

References

[1] Hafner, D., Lillicrap, T., Fischer, I., Villegas, R., Ha, D., Lee, H., & Davidson, J. (2019). Learning latent dynamics for planning from pixels. *36th International Conference on Machine Learning, ICML 2019, 2019-June*.

[2] Hafner, D., Lillicrap, T., Norouzi, M., & Ba, J. (2021). MASTERING ATARI WITH DISCRETE WORLD MODELS. *ICLR 2021 - 9th International Conference on Learning Representations*.

[3] Hafner, D., Pasukonis, J., Ba, J., & Lillicrap, T. (2023). *Mastering Diverse Domains through World Models*.