

The Learning Hypothesis on Spatial Receptive Field Remapping

Henry Kuo (hkuo@college.harvard.edu)

Harvard College, Harvard University
Cambridge, MA 02138 USA

Paul Masset (paul.masset@mcgill.ca)

Department of Psychology, McGill University
Montréal, QC, H3A 1G1, Canada

Blake Bordelon (blake_bordelon@g.harvard.edu)

John Paulson School of Engineering and Applied Sciences,
Center for Brain Science,
Kempner Institute for the Study of Natural & Artificial Intelligence,
Harvard University
Cambridge, MA 02138 USA

Cengiz Pehlevan (cpehlevan@g.harvard.edu)

John Paulson School of Engineering and Applied Sciences,
Center for Brain Science,
Kempner Institute for the Study of Natural & Artificial Intelligence,
Harvard University
Cambridge, MA 02138 USA

Abstract

Grid cells in the medial entorhinal cortex (MEC) and place cells in the hippocampus fire only at specific locations in the environment. Their firing fields are known to change to spatial and nonspatial alterations of the environment. A recent study showed that their receptive fields move toward rewards after rats learn the locations of the rewards. We hypothesize that this change in receptive fields could help animals learn better. To verify this hypothesis, we use receptive fields as features in a simulated agent performing Temporal Difference learning. First, we observed that an evolutionary algorithm would move mathematical place cells closer to rewards to speed up learning. Second, using rat MEC neural data from the study, we found that the simulated agent can learn better with the post-learning receptive fields when evaluating post-learning behavior. These results show that distortions in spatial receptive fields could be a feature of the brain to help animals perform learning tasks.

Keywords: grid cells; place cells; spatial field remapping; temporal difference learning; reinforcement learning

Introduction

Grid and border cells in the medial entorhinal cortex (MEC) and place cells in the hippocampus are known to encode spatial information, firing at specific locations in the environment (Tukker et al., 2022; Moser, Moser, & McNaughton, 2017). The MEC cells have been widely hypothesized to have environment-independent spatial firing patterns, though this has been challenged by recent experiments recording the neurons under more naturalistic conditions, such as in irregularly shaped arenas (Moser et al., 2017; Krupic, Bauza, Burton, Barry, & O’Keefe, 2015). Place cells, on the other hand, are known early on to respond to many spatial and nonspatial cue changes (Wiener, Paul, & Eichenbaum, 1989; Moser et al., 2017). In a recent study, both MEC grid cells and hippocampal CA1 place cells’ fields were found to change during presence of rewards (Boccaro, Nardin, Stella, O’Neill, & Csicsvari, 2019). With these observations, it is hypothesized that MEC cells may instead provide local measures of distance and the distortions in fields may be beneficial for animals to perform tasks (Ginosar, Aljadeff, Las, Derdikman, & Ulanovsky, 2023).

We extend and examine the latter hypothesis using a computational framework to understand if the changes in spatial receptive fields under presence of rewards can be beneficial for animals to learn. In other words, could the animals be changing their representation of the environment in order to learn better? We used artificial place cells (radial basis functions, RBFs) and neural data of MEC cells from (Boccaro et al., 2019) as features in a reinforcement learning framework to understand the computational benefits of spatial receptive field remapping.

Background and Methods

Boccaro et al. (2019) found that cells in the MEC and CA1 region are attracted to goals. In the study, rats were trained to learn three hidden reward locations on a cheeseboard maze while recording from the MEC and CA1. Each trial consisted of three phases: pre-learning, learning, and post-learning. During pre-learning and post-learning, the spatial receptive fields of neurons were recorded in absence of rewards. The study observed that neurons have higher average firing rates near reward locations (Figure 1). We represent the average firing rates of all neurons at location s as $\psi(s)$. The receptive field pre-learning is denoted as ψ_{pre} while that of post-learning is ψ_{pos} .

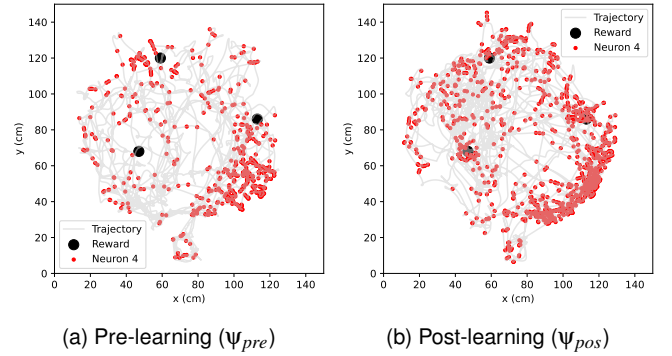


Figure 1: An example MEC neuron’s receptive field before and after the rat learns the reward locations. In the post-learning map, the neuron activates more near the reward locations. The animal also spends more time near reward areas post-learning.

In this paper, we are interested in performing policy evaluation with these receptive fields as features. Consider an environment with the set \mathcal{S} of states and a reward function $R(s)$. The agent follows fixed policy π and seeks to find a value function that measures the expected discounted sum of future rewards starting from state s_0 :

$$v_{\pi}(s_0) = \mathbb{E} [R(s_0) + \gamma R(s_1) + \gamma^2 R(s_2) + \dots] \quad (1)$$

where $0 \leq \gamma \leq 1$ is the discount factor. The agent estimates the value function with $V_{\pi}(s; \mathbf{w}) = \psi(s) \cdot \mathbf{w}$ and tries to minimize the value error:

$$\overline{VE}(\mathbf{w}) \equiv \sum_{s \in \mathcal{S}} [v_{\pi}(s) - V_{\pi}(s; \mathbf{w})]^2. \quad (2)$$

The agent can learn \mathbf{w} by sampling trajectories of π for T steps and applying the Temporal Difference (TD) learning rule (Sutton & Barto, 2018):

$$\mathbf{w}_{n+1} = \mathbf{w}_n + \eta \sum_{t=1}^T [R(s_t) + \gamma V(s_{t+1}) - V(s_t)] \psi(s_t). \quad (3)$$

We also need to model the agent’s policy after the rat’s behaviors before and after learning. Observing that $\pi(s_{t+1}|s_t) =$

$p(s_{t+1}|s_t) = \frac{p(s_{t+1}, s_t)}{p(s_t)}$, we use the triweight kernel density estimator (Marron & Nolan, 1988) to derive the densities of $p(s_{t+1}, s_t)$ and $p(s_t)$, separately. This allowed us to interpolate the data and obtain good estimates of π_{pre} and π_{pos} , the simulated policies of the rat pre- and post-learning.

Furthermore, we are interested in finding how place cell features move to allow fast TD learning. The cross-entropy (CE) based basis adaptation evolutionary algorithm (Menache, Mannor, & Shimkin, 2005) finds features (ψ) that would lower convergence value error in TD learning. It does so by parameterizing $\psi(\cdot; \theta)$ and iteratively find θ s that perform better. In this paper, θ represents the centers of RBFs. In each generation, different θ s are sampled from a Gaussian distribution $\theta \sim \mathcal{N}(\mu, \Sigma)$ and the θ s that have lowest convergence value errors are used to update μ and Σ .

Results

First, we used the CE algorithm (Menache et al., 2005) to see how RBF (place cell) centers would move to increase TD learning speed (Figure 2). We can observe that the closest place cell centers evolved closer to rewards (Figure 2a). Interestingly, the place cells furthest away from the rewards moved further away. We can also confirm that the RBF centers gather near the rewards after the evolution process (Figure 2b). This supports our hypothesis and matches the qualitative observation in (Boccaro et al., 2019) that receptive fields would move closer to rewards.

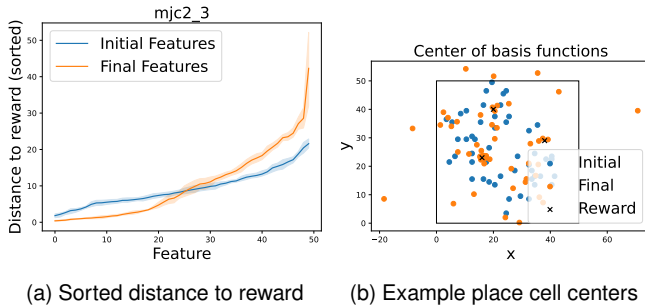
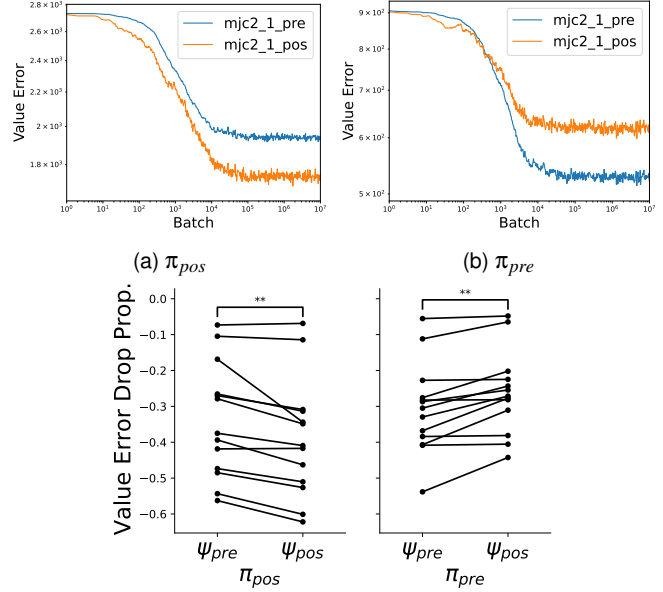


Figure 2: Results from the CE algorithm. (a) Sorted place cell center distance to closest reward location. Shaded area denotes 95% confidence interval over 5 seeds. (b) Example evolution of place cell centers before and after performing the CE algorithm. The box represents the region where the artificial agent could move.

Next, inspired by recent theory of TD learning dynamics (Bordelon, Masset, Kuo, & Pehlevan, 2023), we aimed to see if using the MEC receptive fields from the post-learning phase (ψ_{pos}) would allow a TD learner to learn faster (Figure 3). We found that TD learners following π_{pos} and using ψ_{pos} have faster convergence and lower convergence value errors compared to that of ψ_{pre} (Figure 3a). However, we do not see the same phenomenon when the TD learners follow π_{pre} (Figure 3b). This implies that the improvement in learning arises from the joint alignment of the policy and the representations.



(c) Student t-test on convergence value error drop proportions

Figure 3: (a, b) Representative learning curves of a TD learner using pre-learning and post-learning receptive fields. (a) Learning curves of TD learning evaluating a policy obtained from post-learning rat behavior. (b) Learning curves of TD learning evaluating a policy obtained from pre-learning rat behavior. (c) Value error drop proportion on all 13 trials in the data. The drop proportion is calculated by $(\overline{VE}_{convergence} - \overline{VE}_{initial}) / \overline{VE}_{initial}$. The $\overline{VE}_{convergence}$ is computed by averaging the final 100,000 value errors across 10 seeds. ****P < 0.01.**

Discussion

In this study, we examined whether changes in receptive fields during presence of rewards can be beneficial for animals to learn. Our results support the hypothesis that animals change their receptive fields to learn better.

One limitation of our work is that we only modeled policy evaluation of the learning process. In the complete learning process, the agent not only perform policy evaluation, but also policy iteration. For a complete explanation of the rat's behavior, it would be necessary to model both parts of the learning process.

Another limitation is the low number of features for learning. Modern theory on TD learning (Bordelon et al., 2023) relies on the assumption that features can fully express the target value function. However, we have at most have 45 neurons per trial. This can be solved with new high-throughput neural recording techniques (Steinmetz et al., 2021).

We also recognize that not all cells in the MEC encodes a spatial firing pattern that is tied to the location of the animal. For example, head direction cells in the MEC fires maximally when the animal is facing a certain direction in the environment (Tukker et al., 2022). There are also conjunctive cells that have firing fields similar to a mixture of head and grid cells (Sargolini et al., 2006).

Acknowledgments

HK was supported by the Harvard College Research Program. This work has been made possible in part by a gift from the Chan Zuckerberg Initiative Foundation to establish the Kempner Institute for the Study of Natural and Artificial Intelligence. We thank Jozsef Csicsvari and Charlotte Boccara for kindly providing us with their data, Michele Nardin for help with data curation, and Kiah Hardcastle for useful comments on this paper.

References

- Boccara, C. N., Nardin, M., Stella, F., O'Neill, J., & Csicsvari, J. (2019, March). The entorhinal cognitive map is attracted to goals. *Science*, *363*(6434), 1443–1447. doi: 10.1126/science.aav4837
- Bordelon, B., Masset, P., Kuo, H., & Pehlevan, C. (2023, November). Loss Dynamics of Temporal Difference Reinforcement Learning. In *Thirty-seventh Conference on Neural Information Processing Systems*.
- Ginosar, G., Aljadeff, J., Las, L., Derdikman, D., & Ulanovsky, N. (2023, June). Are grid cells used for navigation? On local metrics, subjective spaces, and black holes. *Neuron*, *111*(12), 1858–1875. doi: 10.1016/j.neuron.2023.03.027
- Krupic, J., Bauza, M., Burton, S., Barry, C., & O'Keefe, J. (2015, February). Grid cell symmetry is shaped by environmental geometry. *Nature*, *518*(7538), 232–235. doi: 10.1038/nature14153
- Marron, J. S., & Nolan, D. (1988, December). Canonical kernels for density estimation. *Statistics & Probability Letters*, *7*(3), 195–199. doi: 10.1016/0167-7152(88)90050-8
- Menache, I., Mannor, S., & Shimkin, N. (2005, February). Basis Function Adaptation in Temporal Difference Reinforcement Learning. *Annals of Operations Research*, *134*(1), 215–238. doi: 10.1007/s10479-005-5732-z
- Moser, E. I., Moser, M.-B., & McNaughton, B. L. (2017, November). Spatial representation in the hippocampal formation: A history. *Nature Neuroscience*, *20*(11), 1448–1464. doi: 10.1038/nn.4653
- Sargolini, F., Fyhn, M., Hafting, T., McNaughton, B. L., Witter, M. P., Moser, M.-B., & Moser, E. I. (2006, May). Conjunctive Representation of Position, Direction, and Velocity in Entorhinal Cortex. *Science*, *312*(5774), 758–762. doi: 10.1126/science.1125572
- Steinmetz, N. A., Aydin, C., Lebedeva, A., Okun, M., Pachitariu, M., Bauza, M., . . . Harris, T. D. (2021, April). Neuropixels 2.0: A miniaturized high-density probe for stable, long-term brain recordings. *Science*, *372*(6539), eabf4588. doi: 10.1126/science.abf4588
- Sutton, R. S., & Barto, A. G. (2018). *Reinforcement learning: An introduction* (Second edition ed.). Cambridge, Massachusetts: The MIT Press.
- Tukker, J. J., Beed, P., Brecht, M., Kempster, R., Moser, E. I., & Schmitz, D. (2022, April). Microcircuits for spatial coding in the medial entorhinal cortex. *Physiological Reviews*, *102*(2), 653–688. doi: 10.1152/physrev.00042.2020
- Wiener, S. I., Paul, C. A., & Eichenbaum, H. (1989, August). Spatial and behavioral correlates of hippocampal neuronal activity. *Journal of Neuroscience*, *9*(8), 2737–2763. doi: 10.1523/JNEUROSCI.09-08-02737.1989