Exploring the Tolman-Eichenbaum Machine for Modeling Spatial Understanding in Robots

Michal Kostrhun, Igor Farkaš

Faculty of Mathematics, Physics and Informatics, Comenius University Bratislava
Mlynská dolina 6284, 842 48 Bratislava
kostrhun1@uniba.sk, igor.farkas@fmph.uniba.sk

Abstract

Spatial understanding is a crucial cognitive component in all agents acting in 3D physical space, including people and robots. We investigate the Tolman-Eichenbaum Machine –a unifying mechanistic framework for the hippocampal role in spatial and non-spatial tasks– for use in biologically inspired Simultaneous Localization and Mapping (SLAM) in robots. Establishing the preliminary context of neuroscience, we provide an approachable overview to the model and its capabilities. We compare it to other approaches used in previous bio-inspired SLAM systems and address their limitations. We summarize advantages and limitations of TEM and briefly outline our future work regarding its use in cognitive robotics.

1 Introduction

Navigation of robotic agents in real-world environments is a complex task that has been actively researched for decades. In particular, we focus on simultaneous localization and mapping (SLAM). The main challenge with SLAM has always been the spatio-temporal complexity of the environment. Navigation in static environments has been computationally solved in the 1990's (Oriolo et al., 1998), however as the environments get more dynamic and complex, traditional methods fall short of achieving efficient solutions. Current visual SLAM methods fail to achieve real-time performance (Sun et al., 2017) and leveraging semantic information to develop semantic SLAM remains an active area of research (Nahavandi et al., 2025).

In our work, we take a different approach. It is known that hippocampus in mammal brains is essential for survey navigation and evidence points towards its involvement in other types of navigation where location memory is essential (Eichenbaum, 2017). The Tolman-Eichenbaum Machine (TEM) is a machine learning-based computational model of hippocampus and entorhinal cortex, formally linking them together (Whittington et al., 2020). The authors demonstrated that TEM cells resemble different cell types also found in organic hippocampus (grid cells, place cells) and learns to path integrate in unknown environments.

In this article we introduce TEM in the context of neuroscience and cognitive science, going over important preliminaries to understanding the model. We further introduce extensions to the original TEM, aiming to make it suitable for use in SLAM tasks, which will also be the subject of our future work.

2 Preliminaries

The hippocampus is a seahorse-shaped structure located in temporal lobe of the brain. It receives input from the entorhinal cortex (Witter, 2011) and plays an essential role in creating cognitive maps of the environment in animals and humans (biological agents). The neuronal representations of cognitive maps are facilitated by various neuronal cell types, including, but not limited to grid cells and place cells.

Grid cells are the most abundant cell type in medial part of EC. They represent a role of a tiling, layered positioning system in the real world. A single grid cell represents a regular, hexagonal grid of fields. When a biological agent enters one of the fields in the real world, the corresponding grid cell starts firing (Rowland et al., 2016).

Place cells are located primarily in the hippocampus. Their firing signals encode a particular location in real-world environment. Due to remapping, a single place cell can encode different locations depending on the environment (Moser et al., 2015).

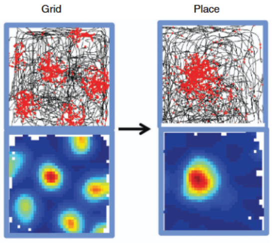

Fig. 1: Grid cell and place cell (Moser et al., 2015). **(Top)** Location of neural spikes (red) superimposed onto trajectory (black) of a biological agent. **(Bottom)** Color-coded rate map with red showing high activity and blue showing low activity.

3 Previous work in Bio-inspired SLAM

Remarkably, the neural representations and processes in the hippocampal-EC system align with the task of SLAM. A biological agent placed in a new environment organizes the perceived stimuli into a cognitive map which is then referenced when localization is performed.

3.1 Overview

We acknowledge two previous works in the field of biologically inspired SLAM models. These works are RatSLAM by Milford et al. (2004) and ViTa-SLAM (Visuo-Tactile SLAM) by Struckmeier et al. (2019). They were chosen as illustrative examples for both older and more recent methods of modeling bio-inspired SLAM.

RatSLAM implements the hippocampal processes using a competitive attractor network of "pose cells" that combine the functions of biological place and head direction cells. By mimicking rodent navigation through the dynamics of "activity packets" and using Hebbian learning for visual association, the system creates a flexible, non-Cartesian representation capable of maintaining multiple spatial hypotheses to resolve environmental ambiguity.

ViTa-SLAM implements the hippocampal-EC model by using a 3D pose cell network designed as a Continuous Attractor Network (CAN) to emulate the spatial representation of grid cells. To resolve environmental ambiguity, the system mimics the biological reliance on whiskers (vibrissae) to supplement coarse vision, fusing unique visual templates with non-unique tactile feedback. This integration utilizes active whisking strategies to enhance sensory data quality and produces a semi-metric "ViTa map", replicating the biological process of organizing disparate stimuli into a cohesive cognitive map for robust localization.

3.2 Limitations

While both of the above works demonstrate novel, bio-inspired SLAM mechanisms for real robots, several technical and practical limitations persist across these models.

Evaluation often focuses on topological consistency rather than Cartesian accuracy, which may not meet standard metric requirements. A primary bottleneck is the high computational cost of maintaining large competitive attractor networks as environments grow. Practically, the reliance on non-standard sensors, such as the use of biomimetic whiskers, limits the immediate portability of these systems to general robotic platforms.

Additionally, correcting accumulated path integration drift upon returning to a previously visited area remains difficult. This phenomenon is predominantly observed in large, circular arenas where topological relationships are harder to resolve.

The biological implementation also remains a highly simplified abstraction. RatSLAM integrates position and orientation into a single network, whereas biological agents utilize distinct populations of place cells and head-direction cells. Finally, these systems lack advanced biological mechanisms like predictive coding and active spatial attention, which are fundamental to how animals filter sensory data and adapt to new environments.

4 Tolman-Eichenbaum Machine

The Tolman-Eichenbaum Machine (TEM), introduced by Whittington et al. (2020) builds upon research advancements in the field of neuroscience. Notably, the idea that relational memory and spatial reasoning may share a common mechanism (Eichenbaum and Cohen, 2014) and that this mechanism can be modeled by finding an efficient representation of a connected graph (Stachenfeld et al., 2017).

Implementing these ideas, authors demonstrate that TEM generalizes structural knowledge in space and non-space (SNS), while also correlating with the neural phenomenons measured empirically in hippocampus of rat and mice. Remarkably, hippocampal remapping of place cells, previously thought to be random, was predicted by TEM to preserve hidden structural representations. These results suggest TEM to be a general framework for hippocampal-entorhinal representation, inference, and generalization across spatial and non-spatial tasks.

4.1 Overview

In the original paper by Whittington et al. (2020), the authors consider an unsupervised learning problem where an agent is moved across the nodes of a sensory graph, receiving each edge transition and sensory input from the current node as an input pair. A sensory graph's edges represent structured relations in SNS, such as the bonds of a family tree (Fig. 2A) or discrete, ordinal movements in euclidean space (Fig. 2B). Each node in a sensory graph represents a unique sensory information.

The TEM is designed to resemble the hippocampal-EC formation. It receives sensory input encoded by lateral entorhinal cells, which is combined with abstract location representation provided by medial entorhinal cells using a conjunctive process in TEM's hippocampus. Hebbian learning on auto-associative networks is used to model TEM's hippocampal place cells. The mechanism behind abstract location in TEM is represented by a recurrent neural network (RNN)-like model.

On a high level, TEM operates by first learning the structure of a family of sensory graphs representing a particular SNS during training. This trains TEM's structural representation, or grid cells. Then, an agent is placed on a sensory graph and TEM, running separately, receives the agent's edge transitions and sensory information from each visited node. On each step, TEM stores its representation of input sensory information ("what") in conjunction with its location representation ("where") inside its auto-associative network. This allows the future retrieval of "what" upon providing "where" (and vice versa) via attractor mechanics. TEM predicts the next "what" by inferring the next "where" from the desired transition, then retrieving the missing "what" of the conjunction with "where" using attractor mechanics. This is, of course, assuming that the node has already been visited.

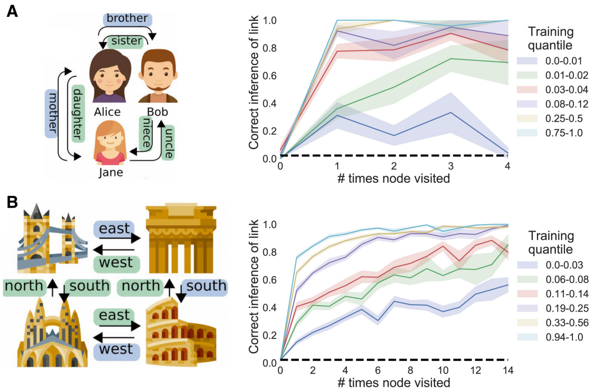

Fig. 2: Learning to learn: when TEM has only seen a few environments (blue/green) it takes many visits to each node to remember it (Whittington et al., 2020). This is because it (1) does not yet understand the structure of the graph and (2) has not learned how to use memories. After visiting more environments and learning the common structure (cyan/yellow), TEM correctly predicts a node on the second visit regardless of the edge taken—TEM now understands both the rules of the graph (path integration) and how to store and retrieve memories. **(A)** social hierarchies, and **(B)** 2D graphs.

4.2 Advantages

Path integration accumulates errors, which the TEM architecture addresses by allowing the refinement of abstract location representations through the retrieval of conjunctive memories. Sensory data acts as a landmark; the TEM indexes its hippocampal memory with current observations to retrieve candidate locations, which better estimate the real location using attractor dynamics. This mechanism provides a more biologically grounded solution for resolving environmental ambiguity than the competitive attractor dynamics used in RatSLAM and its descendants.

Furthermore, as Whittington et al. (2020) demonstrated in their original paper, the TEM architecture empirically resembles the biological hippocampal-entorhinal formation. While RatSLAM simplifies rodent navigation by integrating position and orientation into a single "pose cell" network, TEM maintains distinct populations for structural abstraction (medial entorhinal cortex) and sensory binding (hippocampus). This separation enables the emergence of diverse cell types, including grid, border, and object vector cells that reflect the transition statistics of the environment.

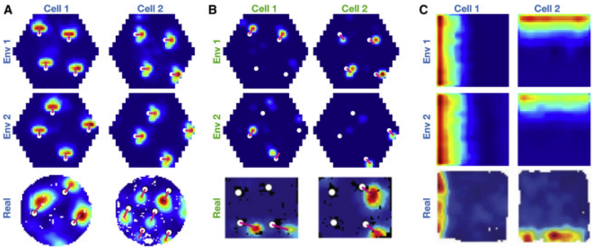

Fig. 3: TEM's various cell types compared to real data measured from rodents (Whittington et al., 2020). **(A)** Object vector cells, which generalize to any object, firing when agent is located at a constant vector offset from an object. **(B)** Landmark cells, similar to object vector cells, though only for specific objects. **(C)** Boundary cells, firing when agent is near a specific boundary.

A primary advantage for robotics is TEM's ability to generalize to novel environments. By factorizing knowledge into structural regularities ("where") and sensory content ("what"), which are separate during training, the system can transfer learned spatial rules to new sensory graphs. Once the underlying structure is learned, a robotic agent can perform first-presentation inferences, predicting transition consequences for nodes it visited through edges it has never physically traversed.

4.3 Limitations

Despite these advantages, several challenges must be addressed to make TEM a viable SLAM solution for general robotics. First, the original TEM architecture is designed to operate on discrete graphs, where movement is defined by edge transitions between adjacent nodes. Real-world robotic navigation occurs in continuous 3D Euclidean space and time, requiring the model to be extended beyond simple ordinal transitions to handle non-discrete coordinates and the full spatio-temporal complexity of robotic tasks.

Second, while TEM works with arbitrary sensory information (often represented as 1-hot vectors in simulations), robots require the processing of structured high-dimensional data, such as vision. While SLAM is most often solved using visual input, the original model's reliance on random sensory values at each node of the sensory graph does not account for the multimodal and complex nature of real-world stimuli. To be effective, TEM must be integrated with robust visual encoders, deriving usable and distinct features from raw pixel data into the representations required for hippocampal binding.

Finally, the computational demands of the architecture may pose a barrier to real-time performance, a limitation already observed in existing visual SLAM methods. The use of iterative attractor networks for memory retrieval and the reliance on training processes like backpropagation through time can be computationally expensive. As environments grow in complexity, maintaining the large auto-associative matrices necessary for high-capacity memory may lead to bottlenecks on mobile robotic platforms.

5 Future work

In the original article, Whittington et al. (2020) introduced the TEM and proved its capabilities in robust mapping and localization across space and non-space. Showing promising results across the various test cases designed to unambiguously test its relational memory, TEM's hippocampal and EC cells have demonstrated empirically similar behavior to real ones found in rodents. Remarkably, TEM's architecture hints at explanations of previously unexplained phenomena, such as remapping of hippocampal place cells.

While the original article focused on proving the above, little, if any, follow-up work has been done to utilize this model in robotics. In our future work, we will do just that. Addressing the above limitations, we will implement the model's architecture in a humanoid robot NICO. We will devise test cases and environments to test the model's capabilities in simulated 3D environments translatable to real-world scenarios. Upon successful implementation, we anticipate the first challenge to be the robot's dynamic peripersonal space where objects can be physically manipulated (by the robot or another agent). Lastly, we will measure our model's performance using standardized metrics and will compare our results with other, bio-inspired and state-of-the-art SLAM systems.

Acknowledgement

This work was supported by Horizon Europe MSCA project no. 101072488. M.K. acknowledges the grant from the Slovak Society for Cognitive Science.

References

Eichenbaum, H. (2017). The role of the hippocampus in navigation is memory. *Journal of Neurophysiology*, 117(4):1785–1796.

Eichenbaum, H. and Cohen, N. (2014). Can we reconcile the declarative memory and spatial navigation views on hippocampal function? *Neuron*, 83(4):764–770.

Milford, M. J., Wyeth, G. F. and Prasser, D. (2004). RatSLAM: a hippocampal model for simultaneous localization and mapping. In *IEEE International Conference on Robotics and Automation*, pp. 403–408.

Moser, M.-B., Rowland, D. C. and Moser, E. I. (2015). Place cells, grid cells, and memory. *Cold Spring Harbor perspectives in biology*, 7(2):a021808.

Nahavandi, S., Alizadehsani, R., Nahavandi, D., Mohamed, S., Mohajer, N., Rokonuzzaman, M. and Hossain, I. (2025). A comprehensive review on autonomous navigation. *ACM Computing Surveys*, 57(9):1–67.

O'Keefe, J. and Nadel, L. (1979). The cognitive map as a hippocampus. *Behavioral and Brain Sciences*, 2(4):520–533.

Oriolo, G., Ulivi, G. and Vendittelli, M. (1998). Real-time map building and navigation for autonomous robots in unknown environments. *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)*, 28(3):316–333.

Rowland, D. C., Roudi, Y., Moser, M.-B. and Moser, E. I. (2016). Ten years of grid cells. *Annual review of neuroscience*, 39(1):19–40.

Stachenfeld, K. L., Botvinick, M. M. and Gershman, S. J. (2017). The hippocampus as a predictive map. *Nature neuroscience*, 20(11):1643–1653.

Struckmeier, O., Tiwari, K., Salman, M., Pearson, M. J. and Kyrki, V. (2019). Vita-SLAM: A bio-inspired visuo-tactile slam for navigation while interacting with aliased environments. arXiv:1906.06422.

Sun, Y., Liu, M. and Meng, M. Q.-H. (2017). Improving RGB-D SLAM in dynamic environments: A motion removal approach. *Robotics and Autonomous Systems*, 89:110–122.

Whittington, J. C., Muller, T. H., Mark, S., Chen, G., Barry, C., Burgess, N. and Behrens, T. E. (2020). The Tolman-Eichenbaum machine: unifying space and relational memory through generalization in the hippocampal formation. *Cell*, 183(5):1249–1263.

Witter, M. (2011). Entorhinal cortex. *Scholarpedia*, 6(10):4380.