Hierarchical Planning in Multilayered State-Action Networks

Matthias Brucklacher^{1,2}, Hanspeter A. Mallot¹ and Tristan Baumann¹

1- University of Tübingen - Cognitive Neuroscience Auf der Morgenstelle 28, 72076 Tübingen - Germany

2- University of Amsterdam - Cognitive and Systems Neuroscience Group Science Park 904, 1090 GE Amsterdam - Netherlands

Abstract. The ability to decompose large tasks into smaller subtasks allows humans to solve complex problems step-by-step. To transfer this ability to an automated system, we propose a spiking neural network inspired by the neurobiological mechanics of spatial cognition to represent space on multiple levels of abstraction. As behavioral experiments suggest that humans integrate spatial knowledge in a graph of places, neurons in the state-action network encode locations while connections between them represent transition actions. In a series of simulation experiments, the influence of hierarchy on planning speed and on the resulting route choice in comparison to single-level models is investigated. We find that the model chooses biased subgoals in line with experiments on human navigation.

1 Introduction

Evidence for the use of hierarchical representations for decomposition of complex tasks is ample and its advantages have frequently been discussed [1, 2, 3]. One approach to model it, is Hierarchical Reinforcement Learning [2, 3] (HRL). There, memorization of longer state-action sequences with known subgoals, socalled "options", is introduced. The issue then shifts to how these options are learned, termed the "option discovery problem" [2]. In this work, a different approach is presented, in which the hierarchical structure of the environment is explicitly represented. Thereby, we address the shortcoming of mechanistic, neural-level models for hierarchical problem-solving. As suggested in previous work on graph-based navigation [4, 5], the state-action network proposed here is a set of environmental states (the neurons/nodes) connected by transition actions (the synapses/edges). It is then hierarchized by a layer of region neurons to represent extended areas, for example districts within a city, an island surrounded by water or even a whole country. Indeed, evidence for neurons encoding larger navigation segments has been found by [6].

Human navigation in regionalized environments underlies two strong biases. In an experimental study by Wiener et al. [1], subjects first explored a virtual environment and were then asked to navigate to a given landmark. The environments were divided into multiple regions either by clear boundaries (a river) or common landmark categories (animals, cars). When choosing among equally long paths to the goal, subjects preferred routes that crossed less region borders. Interestingly, they also showed a strong bias to access the region containing the goal as quickly as possible, neglecting alternative routes of equal length. In ESANN 2021 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Online event, 6-8 October 2021, i6doc.com publ., ISBN 978287587082-7. Available from http://www.i6doc.com/en/.



Fig. 1: Components of the hierarchical state-action network. Left: Each synaptic connection between neurons is linked to a transition action that participates in voting when active. Strong synapses (bold) dominate the voting process and result in motion (red vector on the right) towards the goal marked by a circle. Strengthening of goal-directed connections is a result of the planning process described in the main text. Right: Region neurons (shown enlarged) represent extended areas in space. Each region neuron is bidirectionally connected to each lower-level neuron within its region (connections omitted for clarity).

this paper, a similar setting is used to assess model behavior at region transitions, since these are critical points for hierarchical planning systems. We study interaction of the hierarchical network with a biologically plausible planning mechanism from [7] who applied it to single-level networks that lacked explicit action representation. This planning mechanism can be executed in parallel, in contrast to serial graph searches such as Dijkstra's algorithm in [4]. Simulation experiments firstly focus on the influence of hierarchical structure on planning time, as faster planning would be a major advantage for natural and artificial system, and secondly on route choice.

2 Path-planning in neural state-action networks

Each neuron in the state-action network illustrated in Figure 1 encodes a memorized state in the environment. This can be a visual feature of an object in the environment, for example a landmark. When the agent recognizes this feature, the respective neuron receives a sensory input current. Together with the global inhibitory current, the position of the agent is then encoded in a localized bump of activity [7], a distributed representation. Connections between neurons are bidirectional and represent transition vectors as shown in Figure 1. Since this work focuses on planning, we assume the network structure to be given and refer to [4] for unsupervised learning of state-action networks.

For path-planning, both sensory input and the global inhibitory system are turned off and the neurons corresponding to the goal location are activated. This leads to a wave of activity propagating throughout the network. As the wave passes, goal-directed connections are strengthened [7]. This is achieved through the use of a time-dependent update rule for synaptic weights that requires the use of spiking neurons. As neuron k spikes after neuron j, the connection from k to j, w_{kj} , becomes stronger than the opposing connection from j to k:

$$\frac{dw_{kj}(t)}{dt} = \delta(s_k) \cdot A^- \cdot \left(1 - \exp\left(-\frac{s_j}{\tau_0}\right)\right) \exp\left(-\frac{s_j}{\tau_{STDP}}\right) \tag{1}$$

Here, A^- is a constant regulating the strength of the weight update, s_j and s_k denote the time since the last spike of neuron j and k respectively. $\tau_0 \ll \tau_{STDP}$ and τ_{STDP} are time constants. Evidence for such a mechanism that requires only locally available information has been found for instance by [8]. After planning, sensory input is turned on again and restores the activity bump at the agent's current location (black neurons in Figure 1). The action to execute is then determined by a voting process, in which actions (vectors in squares in Figure 1) are weighted by both synaptic weights and presynaptic neuronal activity. Phrased differently, during navigation the agent follows a synaptic vector field (SVF) assigning a vectorial action to each position.

3 Results

Increased planning speed To test the influence of hierarchical structure upon planning speed, a two-dimensional network of 24×24 neurons is set up. Each neuron is bidirectionally connected to neighbors one (with $w^{P1} = 6$, Figure 1) and two edges ($w^{P2} = 1$) away on the grid. Wavefront propagation in this singlelevel network is compared to hierarchical networks with an additional layer of 2×2 (respectively 3×3 and 4×4) neurons. These region neurons are connected to neighboring region neurons with $w^{RR} = 3$ and to the lower-level with $w^{PR} = 2$. Since it can be assumed that regions overlap, neurons on borders are connected to region neurons of all adjacent regions (cf. Figure 1) with reduced synaptic weights to normalize input from the higher layer.

A wave of spiking activity is initiated by activating the goal neuron in the upper right corner of the network, creating a goal-directed SVF as it propagates (Figures 2a-c). In the hierarchical networks, the wave travels on both levels. As shown in Figure 2d, addition of the second layer drastically speeds up the planning process. Until all neurons are activated, the single-level network takes 55.6 ms, compared to 32.4 ms of the network with 2×2 regions (not shown), 35.6 ms (3×3) and 38.2 ms for the network with 4×4 regions. This equals a reduction of planning time by up to 42%.

Influence of hierarchy on route choice Following the design of an experiment on human navigation by [1], we set up an environment of discrete places arranged on a grid and distributed across two islands interconnected by bridges (Figure 3). As in [1], the agent is placed on a crossroad and must navigate to a given place on the opposing island. To determine the influence of hierarchical network structure, we compare two agents: one with a single-level representation of the

ESANN 2021 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Online event, 6-8 October 2021, i6doc.com publ., ISBN 978287587082-7. Available from http://www.i6doc.com/en/.



Fig. 2: Influence of hierarchical structure on planning time. (a): Activity in a single-level network after 30 ms. (b) and (c): same as (a) for hierarchical networks of different region sizes, as illustrated by black lines. Regions neurons are shown in increased size. (d): Influence of region size on planning speed $(2 \times 2 \text{ corresponds to the largest regions})$.

environment and a second one with an additional region neuron for each of the two islands. Network connectivity follows the previously described simulation.

In the planning phase, the neuron representing the goal is activated. In the single-level network, the resulting activity wave spreads radially from the goal (Figure 3a), whereas in the hierarchical network, it propagates quicker inside the region containing the goal neuron (Figure 3b) and then slows down at the border. Again, wave propagation is faster in the hierarchical network. The resulting SVFs are oriented towards the goal, but in the hierarchical network, vectors within the left region tend to point stronger to the goal region than in the single-level network. Figures 3c) and d) show the resulting trajectories of an agent that is limited to discrete transitions between the places: these are shown as black horizontal or vertical lines whose thickness indicates the frequency with which each step was chosen. The grey band shows the average trajectory calculated from all node-to-node paths. The single-level model chooses paths leading straight to the goal. In contrast, the hierarchical model displays a preference to reach the goal region as quickly as possible as indicated by the bent grey line. Both models reliably reach the goal on an optimal (shortest possible) route: the

ESANN 2021 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Online event, 6-8 October 2021, i6doc.com publ., ISBN 978287587082-7. Available from http://www.i6doc.com/en/.



Fig. 3: Influence of hierarchy on route choice. (a) and (b): Wavefront propagation from the goal marked by the circle in networks without and with hierarchical structure. For the latter, region neurons are shown in increased size. (c) and (d): SVF established by the wave and resulting trajectories.

single-level model in 88% of trials and the hierarchical model in 90%. Note that here all trajectories consisting of rightward and upward steps in the figure have the same total length (are optimal).

4 Discussion

Addition of a hierarchically superordinate layer to the network leads to an increase in planning speed as well as behavioral effects in line with studies on human navigation. Compared to their single-level counterparts, the hierarchical networks took much less time for the planning process. The increase in wave velocity resulted from an activity wave in the higher layer that provided synaptic input to the wavefront on the lower level. Larger regions provide an advantage here (Figure 2d), but it can be assumed that there is an optimal granularity. As regions become larger, the activity wave on the higher level speeds up, and in the extreme case decays before it can provide input to the lower-level wave.

Simulation experiments on navigational behavior of the model show that a virtual agent, equipped with neurons to represent regions, set the closest transition point to the goal region as a subgoal for navigation, disregarding equally short alternatives. This effect was caused by the altered shape of the propagating wavefront: Activation of a region neuron provided additional synaptic input

to all neurons within that region. This input then led to a quick wave traversal of the goal region while transition to the neighboring region was more costly, slowing the wave down at the borders (Figure 3b). From there, it propagated perpendicularly to the borders – in contrast to the wavefront of the single-level networks that spread radially from the goal (Figure 3a). The resulting SVF of the hierarchical network was thus biased towards region transitions and with it the resulting path choices. This observation is in line with behavior of human subjects [1] that preferred to reach the goal region as quickly as possible, too. Here, model behavior is in contrast to HRL. If, in HRL, options are chosen purely based on the reward structure of the environment, all paths of equal length will be selected with equal probability, as can also be expected from breadth-first path searches. Compared to HRL, subgoal selection is much more straightforward in our model, where it is simply a result of altered wavefront propagation, whereas in HRL it leads to the option discovery problem [2]. To allow for taskdomains beyond navigation, states and actions in the present model would have to be generalized similarly to implementations of HRL [3].

5 Conclusion

The proposed model combines a multilayered state-action network with bioinspired planning. When investigating model behavior, we found that it selects subgoals in line with human subjects. In combination with the model's biological plausibility, this poses it as a candidate explanation for hierarchical navigation, although experimental evidence is not conclusive yet. Computationally, the parallel nature of the propagating wavefront in combination with an increased planning speed through hierarchical structure provides a strong advantage over serial graph searches. The network is thus an attractive solution for autonomous vehicles, captures the advantages of hierarchical representations, and can be implemented energy efficiently in spiking neuromorphic hardware.

References

- J. Wiener, H. A. Mallot, 'Fine-to-coarse' route planning and navigation in regionalized environments, Spatial Cognition and Computation, 3(4):331-358, 2003.
- [2] M. M. Botvinick, Y. Niv and A. G. Barto, Hierarchically organized behavior and its neural foundations: a reinforcement learning perspective, *Cognition*, 113(3):262-280, 2009.
- [3] R. S. Sutton, D. Precup and S. Satinder, Between MDPs and semi-MDPs: A framework for temporal abstraction in reinforcement learning, *Artif. Intell.*, 112:181-211, 1999.
- [4] H. A. Mallot, G. A. Ecke and T. Baumann, Dual population coding for path planning in graphs with overlapping place representations. In *Proceedings of the German Conference* on Spatial Cognition 2020, pages 3-17, Springer, 2020.
- [5] B. Kuipers, The spatial semantic hierarchy, Artif. Intell., 119:191-233, 2000.
- [6] J. J. Bos, M. M. Vinck, L. A. van Mourik-Donga, J. C. Jackson, M. W. Witter and C. M. Pennartz, Perirhinal firing patterns are sustained across large spatial segments of the task environment, *Nature Communications*, 8(1):1-12, 2017.
- [7] F. Ponulak and J. J. Hopfield, Rapid, parallel path planning by propagating wavefronts of spiking neural activity, *Front. Comput. Neurosci.*, 7:98, 2013.
- [8] P. D. Roberts and T. K. Leen, Anti-Hebbian spike-timing-dependent plasticity and adaptive sensory processing, *Front. Comput. Neurosci.*, 4:156s, 2010.