Probing the neural representation of space by training recurrent neural networks to perform spatial localization

Christopher J. Cueva (ccueva@gmail.com)
Center for Theoretical Neuroscience, Columbia University

Xue-Xin Wei (weixxpku@gmail.com) *
Department of Statistics, Columbia University

Abstract

It has been well-established that rodent's Entorhinal Cortex (EC) contains a rich set of spatial correlates which are essential for spatial navigation, including grid cells which encode space using tessellating patterns. Although intensely investigated, the mechanisms and functional significance of these spatial representations remain largely mysterious. As a new way to address these questions, we trained recurrent neural networks (RNN) to perform navigation tasks in 2-d arenas based on velocity inputs. Surprisingly, we find that grid-like spatial response patterns emerge in trained networks, along with units that exhibit other spatial correlates, *e.g.*, border cells and band-like cells. All these different functional types of neurons have been observed experimentally. Our results suggest that grid cells, border cells and others as observed in EC may be a natural solution for representing space efficiently given the predominant recurrent connections in the neural circuits.

Keywords: grid cell; spatial navigation; Recurrent Neural Network; neural coding

Introduction

Neural circuits often need to maintain an internal representation of cognitive variables without external stimuli. Here we will focus on spatial navigation, which typically requires the brain to maintain a representation of location and update it according to the animal's movement and landmarks of the environment. Studies in rodent neurophysiology have revealed a rich set of neural correlates of space in Entorhinal Cortex (EC), including grid cells (Fyhn, Molden, Witter, Moser, & Moser, 2004; Hafting, Fyhn, Molden, Moser, & Moser, 2005), along with border cells, band-like cells and others (see Fig. 1a). The study of the neural underpinning of spatial cognition has provided an important window into how high-level cognitive functions are supported by the neural system. How might the spatial navigation task be solved using a network of neurons? Recurrent neural networks (RNN) seem to be particularly useful for these tasks. Indeed, recurrent-based continuous attractor networks have been one of the main types of models proposed for the formation of grid cells, *e.g.,* (McNaughton, Battaglia, Jensen, Moser, & Moser, 2006)). However, these models require fined tuned connectivity patterns, and the evidence of such connectivity patterns has been largely absent.

Here we present a new approach for understanding the spatial representation in EC. Specially, we trained a RNN to perform spatial navigation tasks. We show that training a RNN with biologically relevant constraints naturally gives rise to a variety of spatial response profiles as observed in rodent EC, including grid-like responses. Our result implies that the neural representation in rodent EC may be seen as a natural way for the brain to solve the navigation task efficiently. More generally, it suggests that RNNs can be a powerful tool for understanding the neural mechanisms of certain high-level cognitive functions[†].

Model

Model description

Our network model consists of a set of recurrently connected units ($N = 100$). The dynamics of each unit in the network $u_i(t)$ is governed by:

$$\tau \frac{dx_i(t)}{dt} = -x_i(t) + \sum_{j=1}^{N} W_{ij}^{\text{rec}} u_j(t) + \sum_{k=1}^{N_{\text{in}}} W_{ik}^{\text{in}} I_k(t) + b_i + \xi_i(t) \tag{1}$$

for $i = 1, \ldots, N$. The activity of each unit, $u_i(t)$, is related to the activation of that unit, $x_i(t)$, through a nonlinearity which in this study we take to be $u_i(t) = \tanh(x_i(t))$. Each unit receives input from other units through the recurrent weight matrix W^{rec} and also receives external input, $I(t)$, that enters the network through the weight matrix W^{in}. Each unit has two sources of bias, b_i which is learned and $\xi_i(t)$ which represents noise intrinsic to the network and is taken to be Gaussian with zero mean and constant variance. The inputs to the network were speed and direction. To perform a 2-d navigation task with the RNN, we linearly combine the firing rates of units in the network. The two linear readout neurons, $y_1(t)$ and $y_2(t)$, are given by the following equation:

$$y_j(t) = \sum_{i=1}^{N} W_{ji}^{\text{out}} u_i(t) \tag{2}$$

Training

We optimized the network parameters W^{rec}, W^{in}, b and W^{out} to minimize the squared error between target x- and y-coordinates from a two dimensional navigation task (performed in rectangular and hexagonal environments) and the network outputs generated according to equation (2).

*Corresponding author

[†]This essay is brief summary of a manuscript which is currently under review elsewhere.

Parameters were updated with the Hessian-free algorithm (Martens & Sutskever, 2011). In addition to minimizing the squared error function mentioned above we regularized the input and output weights. Overall, the training aims to minimize a loss function, that consists of the error of the animal, a metabolic cost, and a penalty for large network parameters.

Results and Conclusions

We run simulation experiments in arenas with different boundary shapes. We find that after training, the network can perform localization accurately based on its inputs. To see what kind of representation the RNN has learned to solve this navigation task, we plot individual neurons' mean activity level as a function of the animal's location during spatial exploration.

Perhaps most interestingly, some of the units in the RNN exhibit clear grid-like responses (Fig. 1b,c). They typically exhibit multiple firing fields, with each field having a roughly circular symmetric or ellipse shape. Furthermore, the firing fields are highly structured, *i.e.,*, when combined, are arranged on a regular lattice. Furthermore, the structure of the response lattice depends on the shape of the boundary. Experimentally, it is shown that rodent (medial) EC contains so-called grid cells which exhibit multiple firing fields that lie on a regular grid (Fyhn et al., 2004; Hafting et al., 2005). The grid-like firing patterns in our simulation are reminiscent of the grid cells in rodent. Furthermore, some neurons in the RNN exhibit selectivity to the boundary (Fig. 1b,c). Typically, they only encode a portion of the boundary, e.g. one piece of the wall in a square shaped environment. Such properties are similar to the border cell discovered in rodent EC recently (Solstad, Boccara, Kropff, Moser, & Moser, 2008). Interestingly, there are also neurons in the RNN that exhibit band-like responses (Fig. 1b, c). Experimentally, neurons with periodic-like firing pattern have been recently reported in rodent EC (Krupic, Burgess, & O'Keefe, 2012). Most of the remaining units also exhibit stable spatial responses, but they do not belong to the above categories. These response profiles can exhibit either one large irregular firing field; or multiple circular firing fields, but these firing fields do not show a regular pattern. Experimentally these type of cells have also be observed. In fact, it is recently reported that the non-grid spatial cells constitute a large portion of the neurons in Layer II and III of rodent EC (Diehl, Hon, Leutgeb, & Leutgeb, 2017). The general agreement between the response properties of our model and the neurophysiology provides strong evidence supporting the hypothesis that neurons in rodent EC form an efficient code for representing self-location based on the velocity input.

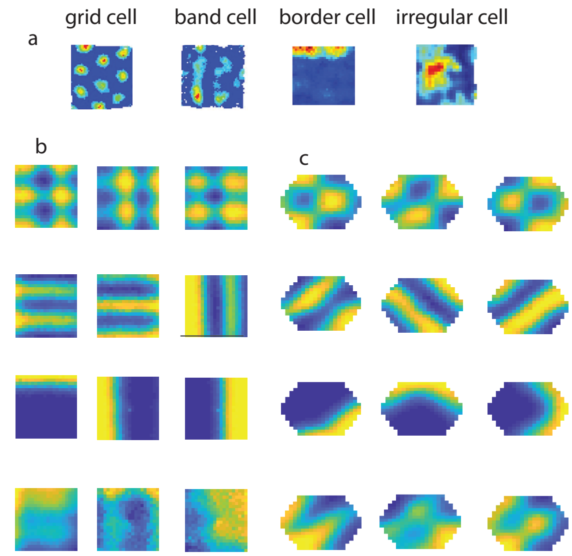

Figure 1: **a** Data from previous studies showing different kinds of neural correlates underlying spatial navigation in rodent EC. All figures are replotted from previous publications. From left to right: a "grid cell" recorded when an animal navigates in 2-d environment, replotted from (Krupic et al., 2012), with the heat map representing the firing rate of this neuron as a function of the animal's location (red corresponds to high firing rate); a "band-like" cell, from (Krupic et al., 2012); a border cell, from (Solstad et al., 2008); an irregular spatially tuned cell, from (Diehl et al., 2017) **b,c** Model responses. Different types of spatial selective responses of units in the trained RNN in square (**b**) and hexagonal (**c**) arenas. Example simulation results for two different environments (square, hexagon) are presented. Blue (yellow) represents low (high) activity. From top to bottom: Grid-like responses; Band-like responses; Border-related responses; Spatially irregular responses.

References

Diehl, G. W., Hon, O. J., Leutgeb, S., & Leutgeb, J. K. (2017). Grid and nongrid cells in medial entorhinal cortex represent spatial location and environmental features with complementary coding schemes. *Neuron*, *94*(1), 83–92.

Fyhn, M., Molden, S., Witter, M. P., Moser, E. I., & Moser, M.-B. (2004). Spatial representation in the entorhinal cortex. *Science*, *305*(5688), 1258–1264.

Hafting, T., Fyhn, M., Molden, S., Moser, M.-B., & Moser, E. I. (2005). Microstructure of a spatial map in the entorhinal cortex. *Nature*, *436*(7052), 801–806.

Krupic, J., Burgess, N., & O'Keefe, J. (2012). Neural representations of location composed of spatially periodic bands. *Science*, *337*(6096), 853–857.

Martens, J., & Sutskever, I. (2011). Learning recurrent neural networks with hessian-free optimization. *ICML*, 1033–1040.

McNaughton, B. L., Battaglia, F. P., Jensen, O., Moser, E. I., & Moser, M.-B. (2006). Path integration and the neural basis of the 'cognitive map'. *Nature Reviews Neuroscience*, *7*(8), 663–678.

Solstad, T., Boccara, C. N., Kropff, E., Moser, M.-B., & Moser, E. I. (2008). Representation of geometric borders in the entorhinal cortex. *Science*, *322*(5909), 1865–1868.