Spatial Representation of Self and Other by Superposition Neural Network Model

Wataru Noguchi[1], Hiroyuki Iizuka[1], Shigeru Taguchi[2] and Masahito Yamamoto[1]

[1]Graduate School of Information Science and Technology, Hokkaido University, Japan
[2]Graduate School of Letters, Hokkaido University, Japan
noguchi@complex.ist.hokudai.ac.jp

Introduction

The sense of self-location is necessary for individuals to perform adaptive behavior like navigation in the environment. The evidence of the sense of self-location is identified as the place cells which was firstly found in the hippocampus of rats' brain (O'Keefe and Dostrovsky, 1971). Recently, it was reported that the locations of other individuals are also represented in the hippocampus of bats and rats (Danjo et al., 2018; Omer et al., 2018). Even some hippocampus cells represent their location without distinction between self and other, and these cells can be considered to be related to empathy mechanisms similar to the mirror neurons (Rizzolatti and Sinigaglia, 2016). However, how such spatial representation that associates self and other's locations is developed is not understood as well as the mirror neurons. One of the popular explanation of representation of the other in self is simulation theory that individuals internally simulate the other's internal recognition. However, the simulation of the other requires another module for inferring the other's internal recognition, which is called model of others. Such simulation module should be different from the module that deal with self internal recognition, and the place cells shared between self and other cannot emerge.

We considered the other's location is represented in the same module as self location rather than in the different module that simulates other's internal recognition. To realize shared representation of self and other's locations like found by (Danjo et al., 2018; Omer et al., 2018), we propose superposition mechanism that two different representations are parallelly processed by the same module at the same time. We implement the superposition mechanism by using two same modules in our proposed neural network model as described later. Through the prediction learning of subjective vision of simulated mobile agent where another agent existed, our proposed network developed shared representation of self and other agent's locations.

Network Model and Simulation

Our proposed network model is constructed to receive subjective vision of self agent $\boldsymbol{v}_t^{self}$ and self motion $\boldsymbol{m}_t^{self}$ and trained to correctly predict future vision $\boldsymbol{v}_{t+1}^{self}$. The network parallelly processes two different representations for self and other through the same module, which we call the superposition module, as follows:

$$\boldsymbol{h}_t^{self} = \phi(\psi^{self}(\boldsymbol{v}_t^{self}), \boldsymbol{m}_t^{self}, \boldsymbol{h}_{t-1}^{self}), \quad (1)$$

$$\boldsymbol{h}_t^{other} = \phi(\psi^{other}(\boldsymbol{v}_t^{self}), \boldsymbol{m}_t^{other}, \boldsymbol{h}_{t-1}^{other}), \quad (2)$$

where ϕ is the function of superposition module, $\boldsymbol{h}_t^{self}$ and $\boldsymbol{h}_t^{other}$ are internal states of the superposition module, and ψ^{self} and ψ^{other} are visual encoders, which are different modules. The processes represented by Eq. (1) and (2) are conducted at the same time, and it is considered the two different states are superposed in the sense that the single superposition module has these two states at the same time. The two superposed states have an explicit difference on the motion input; the self motion $\boldsymbol{m}_t^{self}$ is available, on the other hand, the other's motion $\boldsymbol{m}_t^{other}$ is not available and assumed to be always zero; it means that $\boldsymbol{h}_t^{self}$ becomes self-related representation and $\boldsymbol{h}_t^{other}$ becomes other-related representation. Then, the network generates the prediction of self vision as follows:

$$\boldsymbol{v}_{t+1}^{self} = \hat{\psi}(\boldsymbol{h}_t^{self}, \boldsymbol{h}_t^{other}), \quad (3)$$

where $\hat{\psi}$ is a visual predictor module. In this prediction process, the superposed states are used as two different inputs and integrated for visual prediction. The whole structure of our proposed network is shown in Fig. 1 (a). The superposition module ϕ consists of LSTM (long short-term memory) (Hochreiter and Schmidhuber, 1997), which is an RNN (recurrent neural network) with gate structures, and a fully connected layer as the motion encoder. We implemented the superposition module by using two LSTM modules with same network weights, and we call these two LSTMs the self and other LSTMs for later description, although they are the same module. The visual encoders ψ^{self} and ψ^{other} and visual predictor $\hat{\psi}$ consist of CNN (convolutional neural network).

The network was trained on the visuomotor sensory sequences of a mobile agent in a simulated environment shown in Fig. 1 (b, c). There are two agents as self and other, and four colored boxes as visual landmarks. The self agent

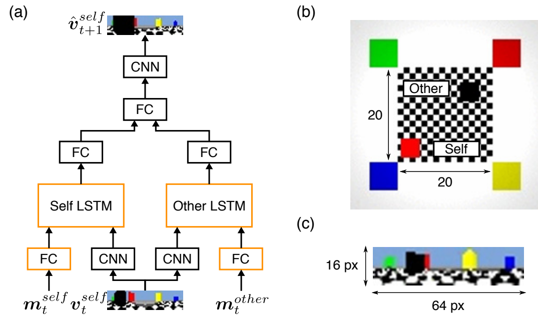

Figure 1: (a) The schematic view of the proposed network model. (b) Overview of the simulation environment. The self agent moves with omni wheels. (c) An example of self agent's vision, which is omni-directional vision.

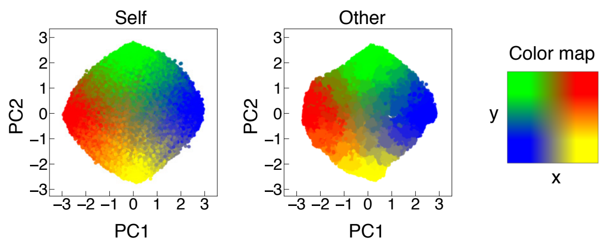

Figure 2: The visualization of the internal states of the self and other LSTMs (left and middle). The states are colored according to the agents' location where colors are assigned for each location as shown in the color map (right).

moves around the environment, on the other hand, the other agent does not move. By relocating the other every trial, the network is trained in various placements of the other agent. We previously showed that the prediction learning of subjective vision realizes self-organization of the representation of self-location in the internal states of an RNN. In the current study, we expected that the development of shared representation of self and other's locations is realized by introducing the superposition module.

Results

After the training, we visualized the internal states of two LSTMs on two-dimensional space by using PCA (principle component analysis) when the agent was moving around (Fig. 2). The visualized internal states are colored according to the self and other agent's locations for self and other LSTM's states, respectively. The same PC space was used for mapping the internal states. It is shown that the internal states are arranged according to actual agents' location and it is considered that the network developed the representation of self and other locations. It should be noted that the self and other's locations seemed to be represented on the same region of the internal state space of self and other LSTMs. Then, we constructed a linear regression model that predict the location of the self agent from the self LSTM's internal states $\boldsymbol{h}_t^{self}$; then predict the location of the other agent from the other LSTM's internal states $\boldsymbol{h}_t^{other}$ using the model constructed for predicting self agent's location. As a result, the error distance for self agent's location became 0.40 and that for other agent's location became 1.38. The error is small considering that theoretical expected distance between two random points sampled from a square size of 20×20 is about 10.4, and it is considered that the superposition module represented the self and other's locations in a shared representation. There are no constraints to make the correspondence of the locations and internal states the same; however, our results show that shared representation of the self and other's location can be developed in the network where the same module represent two different states at the same time.

Discussion

Although it is possible to develop representations shared between two different modules, i.e., not a single superposition module, some learning mechanisms or constraints are necessary to have shared representation for the different modules in addition to developing the representation of self and other's location. On the other hand, our proposed network processes self and other's representation in the same way. In such superposition network, the other can be recognized by applying the self model to the other without constructing the other's model. Such superposing of self and other's model is more efficient than separate development of self and other's model; consequently, the shared representations of self and other's location were developed without any additional constraints. Although such superposition structure with two separate modules with same weights in our implementation has not been found in biological systems, we considered that parallel processing of the self and other's representation in the same network module is necessary for developing the shared representation of the self and other's locations.

Acknowledgements

The authors would like to thank Prof. Masanori Takezawa and Prof. Kenji Ogawa for useful discussions. This work was supported by JSPS KAKENHI Grant Number JP18J20404.

References

Danjo, T., Toyoizumi, T., and Fujisawa, S. (2018). Spatial representations of self and other in the hippocampus. *Science*, 359(6372):213–218.

Hochreiter, S. and Schmidhuber, J. (1997). Long short-term memory. *Neural Comput.*, 9(8):1735–1780.

O'Keefe, J. and Dostrovsky, J. (1971). The hippocampus as a spatial map. preliminary evidence from unit activity in the freely-moving rat. *Brain research*, 34(1):171–175.

Omer, D. B., Maimon, S. R., Las, L., and Ulanovsky, N. (2018). Social place-cells in the bat hippocampus. *Science*, 359(6372):218–224.

Rizzolatti, G. and Sinigaglia, C. (2016). The mirror mechanism: a basic principle of brain function. *Nature Reviews Neuroscience*, 17(12):757.