Presynaptic inhibition and incremental learning in the striatum of the basal ganglia

Hiroyuki Ohta*, Yasuhiro Nishida*, James C. Houk**
*Department of Physiology
National Defense Medical College
ohta@ndmc.ac.jp, ynishida@ndmc.ac.jp
**Feinberg School of Medicine
Northwestern University
j-houk@northwestern.edu

In order to adapt to diverse situations, it is necessary to incorporate new information into the existing model of the world, while keeping previously acquired knowledge intact. We call this process "incremental learning". In order to realize incremental learning, the state expression, which is a part of state-transition, should be in parallel. Conditional and competitive lateral inhibition has to be employed to satisfy the parallelism. Since the striatum in basal ganglia receives parallel inputs from most of the cerebral cortex, the striatum is thought to be the mediator of discovering order relationships between cortical areas. The medium spiny neurons (MSNs) in striatum are postulated to have reciprocal pre-synaptic inhibitory connections, and consequently the model striatum contributes to learning of the sequential activations of cerebral cortical areas in an incremental fashion.

1 Introduction

In this paper, we indicate that the striatum in basal ganglia is suitable for realizing incremental learning, which is to merge a newly added sequence of sensory-motor coupling to the old sequence.

How do human and animals adapt to the new situations without losing the past experience? By categorizing the newly obtained information into the existing classification scheme such as developed by self-organizing map [1], it is possible to update only the corresponding part for new information in the map. However, human and animals need to adapt to completely new situations only by using intermittent and fragmented reward or punishment. In that case, it is impossible to determine what the newly obtained information is, using only the result of evaluation and the ambiguous recognition. Although the generalization capability of neural network may be able to solve a certain part of such classification problem, we should rather consider the partial adaptation of existing pathways in the network without complete classification.

For the last several decades, physiological research on learning in the basal ganglia has advanced our understanding of learning by intermittent and fragmented. On the other hand, the conceptual research of artificial neural network (ANN) has also been done since McCulloch-Pitts type neuron model [2] was proposed. In order to interface the microscopic physiological view about the basal ganglia and the macroscopic behavioral view, the result of ANN research had been referred. However, the general ANN has a conceptual misunderstanding of parallelism in the network and therefore it fails to provide understanding of the real brain. Furthermore, the ANN has still been confronted with the fatal problem, the catastrophic interference [3][4]. Therefore, we reveal the conceptual and practical problem in ANN, provide the alternative view about the parallelism and discuss the functional role of basal ganglia in incremental learning.

2 Catastrophic Interference

In the general ANN, the process of learning a new set of patterns suddenly and completely erased a network's knowledge of what it had already learned [3][4]. This is called "catastrophic interference". Although a lot of neural network models have been proposed, the problem of catastrophic interference has not been addressed sufficiently.

To explain catastrophic interference, we take the recurrent neural network (RNN) with backpropagation through time (BPTT) learning rule [5] as an example. BPTT is an adapted version of standard backpropagation [6] for RNN. The key idea to BPTT is to unfold the RNN into a multilayer feedforward network for adapting backpropagation each time a sequence is processed. Backpropagation generates an error value for layers feeding into the outputs, based on each weight. These error values are propagated backward to another layer and so on, and then the weights in each layer are modified to adapt these error values to zero. Normally, BPTT truncates the continuous input–output sequence to length n, which defines the number of times to unfold and is limited by the size of buffer memory available to train the unfolded layers. This means that BPTT cannot adjust sequences before n time steps and that the new teaching signal overrides the previous sequences before n time steps. Moreover, since the override is broadened to all of the weights in all of the unfolded layers by BPTT algorithms, the previous sequences are immediately broken as soon as the new sequence is added.

In real life when adapting to new situation, although we may sometimes forget the old knowledge, we can usually review and reconstruct existing knowledge to fit to the new situation. Without reconstructing the existing knowledge, we probably cannot create a new solution. In order to understand our ability to exploit the implicit knowledge effectively for new situation, we should focus on the representation form of events in the brain.

3 Rethink Distributed Representation

In the research area of artificial neural network (ANN), although the distributed representation [7] is thought to be the most important property of neural network, the relationship between parallelism and the distributed representation should be reviewed to understand the incremental learning in real brain and the catastrophic interference in ANN. The question whether multiple cells or a single cell should be used for expressing an idea is the pseudo problem. We explain why the question is pseudo problem and how the pseudo problem has been created confusion, for example, the catastrophic interference in neural networks.

3-1 What should be distributed?

The concept of distributed representation was originally proposed as a method for expressing an idea by multiple cells not by a single grandmother cell [7]. Most connectionists have claimed that a single grandmother cell is not appropriate for expressing an idea since a single cell cannot include any rich and varied contents (ex. the qualia of grandmother). However, we can find a paradox in the claim that “an idea” has multiple contents. Expressing multiple contents, the media with multiple nodes must be used. Therefore, the question whether multiple cells or a single cell should be employed for expressing an idea is the pseudo problem.

Although the connectionists have criticized the artificial intelligence (AI) research, connectionism shares the same idea as AI. In the research of AI, the structure of “knowledge” has been variously formulated. The same applies to connectionism as described above. It is hard to avoid our tendency to search the media of an idea. From the external viewpoint, however, our nervous systems receive the parallel signals from sensory neurons and excite the motor neurons in parallel. Each sensory neuron obviously has a unique modality and each motor

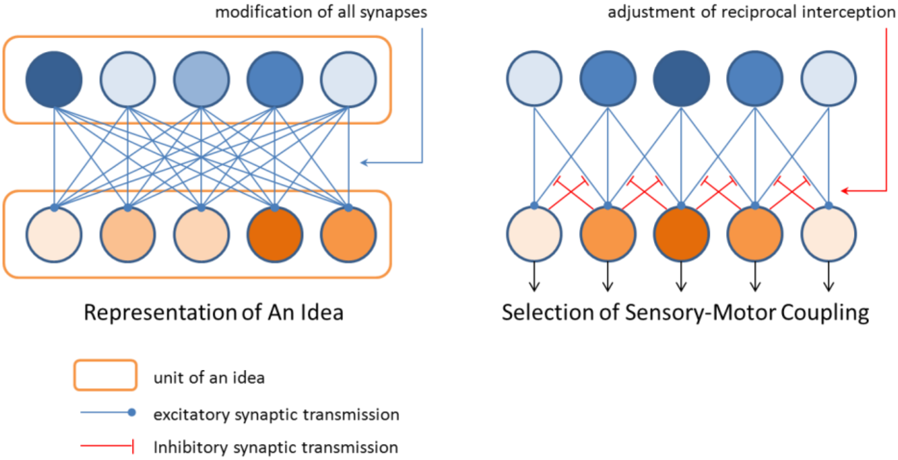

Figure 1. From representation of an idea to selection of sensory-motor coupling.

neuron has a role to contract specific muscles. In sum, each sensory-motor pathway obviously has unique properties respectively. Then, there is no need to adhere to expressing unique ideas and knowledge in order to research behavior learning. Instead, we should investigate how to form sequences of sensory-motor coupling without explicit expression of state-transition.

3-2 From many for one to one for many

We propose the question of how many independent pathways are realized in a neural network, in place of the question of how many neurons participate in the expression of an idea. Implementing the distributed representation in ANN, entire sets of neurons in a layer are recruited for representing an idea (see Fig. 1 left). In that case, however, all of the synapse weights of all of the neurons in a layer suffer from disordered modification during learning, and thereby the catastrophic interference of representation occurs. Although there was an approach to avoid catastrophic interference by using prior categorization [8], it is firstly necessary to consider the mechanism to comprise multiple pathways independently.

In order to comprise multiple independent pathways in a network, the following couple of requirements should be met. First is (i) to preserve the parallel excitatory pathways as much as possible by using competitive mechanisms, such as combination of lateral inhibition, conservation of synaptic weights [9] and hysteresis of up-down state transition [10]. Second is (ii) to modulate the parallel excitatory pathways transiently by using conditional inhibitory transmissions (see Fig. 1 right) in order to adapt the network to the situation immediately without disruption of existing excitatory pathways [11][12][13][14]. Combining those requirements, the selector function and the state expression realized by the neural network will be implicit, and also multiple pathways will share the same neuron on those routes (ref. Dynamical Cell Assembly [15]).

Put simply, the problem is which is better to use in the selector function of neural network: explicit/serial expression or implicit/parallel expression. Although the implicit and parallel expression is difficult to treat especially from the view of mathematical formulation and its realization will not be so mathematically beautiful, we have to retrieve parallelism from the misperception that originally distributed existences were serialized into the oneness.

4 Incremental Learning in Basal Ganglia

We propose a hypothesis that the striatum contributes to the realization of incremental learning of activation order in cerebral cortex areas, since the striatum satisfies the above couple of requirements: (i) the preservation of parallel excitatory pathways and (ii) the modulation of parallel excitatory pathways transiently by using conditional inhibitory transmissions.

The striatum receives inputs from various cerebral cortical areas. Assuming that the topologic relationship between the cortical areas and the striatum is preserved, the parallelism is also satisfied in the striatum obviously. The loop, which is composed by the excitatory pathways from cerebral cortex to striatum, double inhibitory pathways from striatum to thalamus and the excitatory recurrent loop between thalamus and cerebral cortex, contains parallel pathways (see Fig.2 which is modification of Fig.1 in [16] and Fig.4 in [17]). The striatum in the loop will be able to realize the selection of sensory-motor coupling [16][17], instead of expressing an idea by all of neurons on the layer.

And also, since the MSNs in striatum have lateral presynaptic inhibitory transmissions [18][19] and a binary logic-like firing property with hysteresis [10], the MSNs will be able to compete with each other with only a few winners emerging [20]. Modifying the competition between MSNs by some dopaminergic mechanisms [10], the order relation between excitatory parallel pathways from cerebral cortexes to thalamus will be modified transiently.

In conclusion, it is likely that the striatum contributes to the realization of incremental learning of activation order in the cerebral cortex areas.

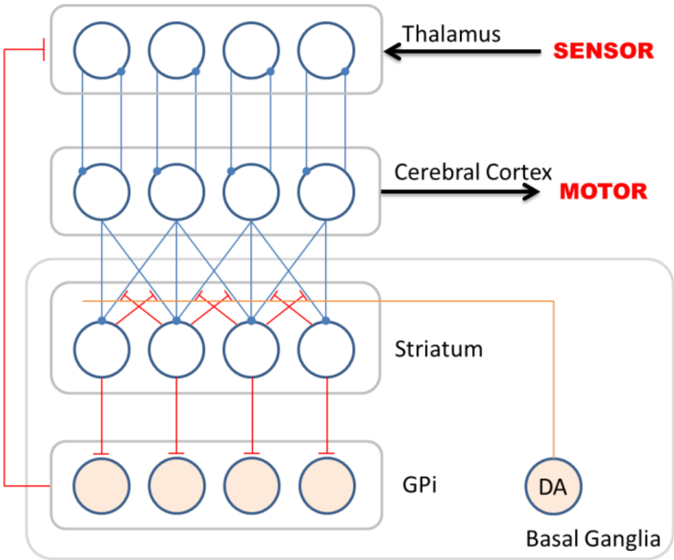

Figure 2. Selection of sensory-motor coupling in striatum through by modification of competitive inhibitions between medium spiny neurons (MSNs) in striatum by dopaminergic input.

Bibliography

[1] Kohonen, T., 1998, Neurocomputing, 21, 1.
[2] McCulloch, W. S. and Pitts, W. H., 1943, Bulletin of Mathematical Biology, 5,115.
[3] McCloskey, M. and Cohen, N.J., 1989, In the psychology of learning and motivation (Bower, G.H., ed.) , 109, Academic Press.
[4] Ratcliff, R., 1990, Psychological Review, 97(2), 285.
[5] Williams, R. J., and Zipser, D. 1990, Backpropagation: Theory, architectures, and applications (Chauvin ,Y., and Rumelhart , D. E., ed), Psychology Press (Hillsdale).
[6] Rumelhart, D. E., Hinton, G. E., and Williams, R. J., 1986, Parallel distributed processing: explorations in the microstructure of cognition, vol. 1, 318, MIT Press (Cambridge).
[7] Hinton, G., McClelland, J. and Rumelhart, D., 1986, Distributed representations, in Parallel Distributed Processing: Explorations in the Microstructure of Cognition: Vol. 1. Foundations, MIT Press (Cambridge).
[8] Polikar, R., Upda, L., Upda, S.S. and Honavar, V., 2001, IEEE Transactions on Systems, Man, and Cybernetics, Part C: Applications and Reviews,31(4) , 497-508.
[9] Royer, S., and Pare, D., 2003, Nature, 422, 518–522.
[10] Gruber, A. J., Solla, S. A., Surmeier, D. J. and Houk, J. C., 2003, J. Neurophysiology, 90, 1095-1114.
[11] Young, J. Z. 1964, A model of the brain. Oxford (Clarendon).
[12] Young, J. Z. 1965, Proceedings of the Royal Society, 163, 285–320.
[13] Young, J. Z. 1966, The memory system of the brain. University of

California Press (Berkeley).
[14] Ohta, H. and Gunji, P-Y., 2006, Neural Networks, 19(8),1106–1119.
[15] Hebb D.O., 1949, The Organization of Behavior: a Neuropsychological Theory, Wiley (New York).
[16] Beiser, D.G. and Houk, J.C., 1998, Journal of Neurophysiology, 79, 3168-3188.
[17] Houk, J.C., Bastianen, C., Fansler, D., Fishbach, A., Fraser, D., Reber, P.J., Roy, S.A and Simo, L.S, 2007, Philosophical Transaction of the Royal Society B, 362, 1573-1583.
[18] Calabresi, P., Mercuri, N. B, De Murtas, M. and Bernardi, G., 1991, Journal of Physiology, 440, 581-599.
[19] Lacey, C. J., Boyes, J., Gerlach, O., Chen, L., Magill, P. J. and Bolam, J. P., 2005, Neuroscience, 136, 1083–1095.
[20] Houk, J. C., Davis, J. L. and Beiser, D. G. eds., 1995, Models of Information Processing in the Basal Ganglia. MIT Press (Cambridge).