(Extended Abstract)

Spike timing dependent plasticity (STDP) to make robot navigation intelligent

Akira Imada

Brest State Technical University
Moskowskaja 267, Brest 224017 Republic of Belarus
akira@bsty.by

Introduction

This talk presents the guidelines of our ongoing project regarding machine intelligence. Our aim is to make an artificial agent act more intelligently. To be more specific, we want the agent to behave differently even when the agent encounters an identical environment to the one it learned before. Let us consider path-planning problem. A-star algorithm is one of the standard approaches to it, and the algorithm is already efficient to search for the shortest route to the goal. Or genetic algorithm, enforcement learning, neural network, whatsoever, can easily find it. What we want, however, is not a deterministic solution, but an artificial agent's *every-time-a-different-action-in-a-same-situation.* To seek this goal, we attack here this path-planning problem using a spiking neural network which learns, during an exploration, how it will modify its synaptic weights. We plan it by means of spiking-timing-dependent-plasticity which might be called a spiking neuron's version of Hebbian learning.

What should be like Machine Intelligence?

Assume, for example, we are in a foreign country where we are not so conversant in its native language, and assume we ask, "Pardon?" or something like that, to show we have failed to understand what they were telling us. Then intelligent people might try to change the expression with using easier words so that we understand this time, while others, perhaps not so intelligent, would repeat the same expression with a little bigger voice.

What if your canary stops singing? In Japan, we have legendary three different strategies for this: (i) *Wait until she sings again*; (ii) *Do something so that she sings again*; and (iii) *Kill her if she doesn't sing any more.* A good suggestion to be intelligent, however, might be *"Be always flexible. Don't stick to one strategy even if you encounter a similar event you met before."*

The point is, we human do not usually behave in the same way as before even when we come across the identical situation again. Then what about machine intelligence?

Though I am working in a department called *"Intelligent Information Technology,"* our research results have not seemed to be very intelligent so far from that point of view. Or, we organize, once in a while, a conference named *"Neural Network and Artificial Intelligence."* Unfortunately, however, even seemingly success reports are not so intelligent from that point of view.

We now are challenging this issue. Specifically we exploit a spiking neural network, also seeking its biological plausibility. Despite of a tremendously lot of elegant realizations of artificial intelligence by spiking neurons, usually those agents with an artificial spiking neural network repeat a series of same actions under an identical situation.

This project was motivated by Floreano's *"Evolution of Learning "* (Floreano et al. 2000) in which a physical micro robot explores a physical world. The robot is controlled by a neural network, though it was not by spiking neurons at least at that time. Starting with a random configuration of synaptic weight values, it learns how each of those synaptic weights are modified during an exploration. It evolves to learn, from one step to the next, *which synapse should use which rule with what parameter values* in order for the robot to move efficiently. Floreano et al. proposed four Hebbian type learning rules for the purpose. Later, Stanley et al. (2003) united these four rules into one equation with two parameters with the meaning being remain intact, so that evolution is just on a pair of these two parameters assigned to each of synapses.

A Benchmark

We propose a benchmark as follows. *In a huge gridworld, agent moves to the neighboring cell spending one unit of energy. Starting from the base located in the center of the gridworld with N units of energy, an agent must travel visiting as many different cells as possible, and the agent must go back to the base before consuming all the energy.* We call it a *Planet Land-rover Problem.*

Model and Method

Amang others, we use an *Integrate-and-Fire* model of spiking neurons following Florian (2007) in which dynamics of neurons is according to:

$$u_i(t) = u_r + \{u_i(t-\delta t) - u_r\}\exp(-\delta t/\tau) + \sum_j w_{ij} f_j(t-\delta t),$$

assuming we simulate the dynamics in discrete time with a time step δt for the sake of simplicity, where $u_i(t)$ is membrane potential of the i-th neuron at time t, u_r is resting potential, τ is decay time constant, and $f_j(t)$ is 1 if j-th neuron has fired at time t or 0 otherwise. When membrane potential exceeds firing threshold θ it is reset to the reset potential which is equal to the resting potential u_r here.[1]

As for the learning of the synaptic weight values, Florian applies:

$$w_{ij}(t+\delta t) = w_{ij}(t) + \gamma r(t+\delta t)\zeta_{ij}(t),$$

where $r(t)$ is reward at time t and γ is *discount rate* by which eventual reward is estimated as

$$r(t) + \gamma r(t+\delta t) + \gamma^2 r(t+2\delta t) + \gamma^3 r(t+3\delta t) + \cdots.$$

Dynamics of ζ_{ij} is given by:

$$\zeta_{ij}(t) = P_{ij}^{+}(t) f_i(t) + P_{ij}^{-}(t) f_j(t),$$

and $P_{ij}^{\pm}$ are:

$$P_{ij}^{+}(t) = P_{ij}^{+}(t-\delta t)\exp(-\delta t/\tau_+) + A_+ f_j(t),$$

$$P_{ij}^{-}(t) = P_{ij}^{-}(t-\delta t)\exp(-\delta t/\tau_-) + A_- f_i(t),$$

where $\tau_\pm$ and $A_\pm$ are constant parameters.[2] See (Florian, 2007) for more in detail.

---

[1] Florian's setting is u_r = -70 mV, θ = -54 mV, τ = 20 ms and δt = 1 ms.

[2] For his benchmark of XOR, he set $\tau_+ = \tau_- = 20ms$, $A_+ = 1$, and $A_- = -1$. P_{ij}^{+} and P_{ij}^{-} track the influence of pre-synaptic and post-synaptic spikes, respectively (Florian 2007).

To be more challenging

The benchmark mentioned above was taken its idea from so-called a *Jeep-Problem*[3] in which an agent navigates a jeep in a 1-D desert with a limited capacity of fuel tank, starting from its base where robot can return later to refill the fuel. The jeep also has a container to put some of its fuels somewhere in the desert for a future usage. The agents repeats the procedure – (i) start the base; (ii) navigate the desert; (iii) put fuels somewhere or get fuels found if necessary; and (iv) return to the base. The goal is thus to maximize its penetration into the desert with n, number of returns to the base, being a parameter.

We extend it to 2-D gridworld. Furthermore, instead of aiming maximum penetration into the desert from the base, we changed it into the exploration starting from the base and returning to the base again. As it is too demanding to begin with, we changed the scenario by setting $n = 1$ in this talk. As a future step we will set it to $n > 1$. It would be a good challenge.

Concluding Remarks

We already have a toy robot like SONY's AIBO. It splendidly learns an environment. It acts differently in a different situation according to how it learned these situations. However, it acts exactly in the same way if it comes across the same situation it has already learned. AIBO can now plays a roll of a wonderful pet, for example. However, this *identical-action-in-identical-situation* would lose the owners interest, sooner or later. From this perspective, I hope today's talk will be a trigger to design a more human like intelligent robot. Also hope to have a stimulating discussion in the seminar expecting a future collaboration.

Acknowledgment

We greatly thank Saulius Maskeliunas for giving us this opportunity to talk in this seminar. It was a good chance for us to get organized what we are thinking of, and what will be the next step for this project.

Reference

Floreano, D., and J. Urzelai (2000) "Evolutionary robots with online self-organization and behavioral fitness." Neural Networks Vol. 13, pp. 431–434.

Stanley, K. O., B. D. Bryant, and R. Miikkulainen (2003) "Evolving adaptive neural networks with and without adaptive synapses." Proceedings of the IEEE Congress on Evolutionary Computation, pp. 2557–2564.

Florian, R. V. (2007) "Reinforcement learning through modulation of spike-timing-dependent synaptic plasticity." Neural Computing, Vol. 19, No. 6. pp. 1468-1502.

[3] http://mathworld.wolfram.com/JeepProblem.html