Eberhard Karls Universität Tübingen
Mathematisch-Naturwissenschaftliche Fakultät
Wilhelm-Schickard-Institut für Informatik

Master Thesis Bioinformatics

Title of thesis

Vor- und Nachname

Datum

Reviewers

Name Erstgutachter
(Bioinformatik)
Wilhelm-Schickard-Institut für Informatik
Universität Tübingen

Name Zweitgutachter
(Biologie/Medizin)
Medizinische Fakultät
Universität Tübingen

Abstract

Write here your abstract.

Zusammenfassung

Bei einer englischen Masterarbeit muss zusätzlich eine deutsche Zusammenfassung verfasst werden.

Acknowledgements

Write here your acknowledgements.

iv

Inhaltsverzeichnis

Abbildungsverzeichnis

Tabellenverzeichnis

Kapitel 1

Methodology

1.1   motor learning model

This work optimizes a motor spicing model introduced by Spüler et al. [SNR15] which is an extension of the work of Chadderdon et al. [CNKL12]. The model presents a biologically realistic model of the motor cortex, which uses reinforcement learning to reach a target angel with a simulated forearm. As a standard reinforcement model [KLM96][CNKL12] it consists of an actor, that maps perceptions to actions a environment that reacts and a critic providing reward or punishment feedback to the actor. Spüler et al. and Chadderton et al. realized the environment as a forearm model with a one degree of freedom joint. Therefore the forearm could be moved either up or down. The joint angle limited by 0°and 135°. Whereby 0°means fully straightened and 135°means fully flexed. The actor got implemented with a spiking neural network. A spiking neural network is a weighted connected graph with some input and output nodes. Each edge is called a connection and each vertex called a neuron. The network calculates a time depending mapping from some input to output values. Each neuron in this model represents an either excitatory or inhibitory dynamic unit. Excitatory neurons excite connected neurons, while inhibitory neurons inhibit connected neurons. The dynamics or in other words the time depending state of each neuron is represented by the membrane potential $V_m(t)$. $V_m(t)$ is modelled by an differential equation introduced by Izhikevich [Izh03], which is a good and efficient approximation of the Hodgkin-Huxley model [HH52]. Izhikevich's model slightly adapted by Spüler et al. with noise is described as

follows (variable names changed):

$$V_m(t)' = 0.04V_m(t)^2 + 5V_m(t) + 140 - u + I(t) + V_n(t)$$
$$u(t)' = a(bV_m(t) - u(t))$$
$$\text{reset after spike:}$$
$$\text{if } v \geq V_t, \text{ then } \begin{cases} 1 & \leftarrow V_r \\ 0 & \leftarrow u + d \end{cases}$$

where u represents the membrane recovery variable. I defines the synaptic input currents. That is the sum of weighted inputs to a neuron.(TODO ref to ANN fundamentals) The firing pattern of a neuron depend on the choice of parameters a, b, d, V_r and V_t. Different resulting spiking patterns can be seen in appendix TODO refpicture. V_r is the resting potential of the neuron. V_t the spiking threshold. V_n is a 300Hz noise input that leads to motor babbling. That is, network send output signals although no input signal is given. This behaviour is important for reinforcement learning since if the actor doesn't produce any actions the critic isn't able to give feedback [CNKL12].The parameter a,b,c,d are chosen as one can shown in table 1.1.

Tabelle 1.1: Parameters used in the dynamic model

Parameter	Excitatory	Inhibitory
a	0.02	$0.02 + 0.08r_i$
b	0.2	$0.25 - 0.05r_i$
d	$8 - 6 \cdot r_i^2$	2
V_r	-65	-63

r_i is uniform distributed in $[0, 1]$. For excitatory neurons moving r_i from 0 to 1 leads to a transition from a regular to a chattering spiking pattern. For inhibitory neurons from a fast to a low-threshold spiking pattern. Examples of those patterns can be seen in (TODO ref appendix.). The neurons are further divided into 5 logical groups: proprioceptive (P), excitatory sensory (ES), inhibitory sensory(IS), excitatory motory (EM) and excitatory sensory (IS) neurons. Spüler's et al. model contains 48 P, 96 ES, 32 IS, 48 EM and 32 IM cells. Which leads to a total amount of 256 neurons. Each logical group is connected with a specific probability as shown in 1.2:

Based on these connections the resulting network is shown in figure (TODO figure network)

The motor control cycle works as follows: A new movement angle is encoded by the EM neurons, after a time gap of 50 ms the forearm moves. At the same time reward or punishment is send. After a second time step of 25 ms the new position angle is encoded by the proprioceptive cells. [SNR15] grounds the time gaps on peripheral and sub cortical processing delays. Why there is an

Tabelle 1.2: Connection probabilities of neuron types in the model of Spüler's et al.

	P	EM	IM	ES	IS
P	0	0	0	0,1	0
EM	0	0	0,43	0	0
IM	0	0,44	0,62	0	0
ES	0	0,08	0	0	0,43
IS	0	0	0	0,44	0,62

additional delay for the P cells to fire but not for the reward keeps ungrounded. The whole system works with a frequents of 0.02kHz. That means the EM neurons encode a new angle every 50ms. An overview of this process can be seen in figure ref TODO. The time needed by EM cells to encode a new angle is again 50ms. EM cells are parted into two equal sized groups of 24 neurons. The first 24 are in the first group the next 24 in the second. Each firing of a cell in the first group leads to an 1°downward motion of the arm; in the lower part it leads to an 1°upward motion. Over the time window of 50ms all spikes in each group are summed up. The resulting movement angle the difference of those sums. Whether reward or punishment is send by the critic is decided whether the distance $\Delta\theta_t$of the current angle θ_t to the target angle θ_{target} got smaller or higher compared to the mean of the last two differences $\Delta\theta_{prev}$. This can be seen in the following formula:

$$\Delta\theta_t = |\theta_t - \theta_{target}|$$

$$\Delta\theta_{prev} = \left|\frac{\theta_{t-1} + \theta_{t-2}}{2} - \theta_{target}\right|$$

$$\text{critic response} = \begin{cases} \text{reward} & \text{if } \Delta\theta_t < \Delta\theta_{prev} \\ \text{punishment} & \text{if } \Delta\theta_t > \Delta\theta_{prev} \\ \text{no response} & \text{if } \Delta\theta_t = \Delta\theta_{prev} \end{cases}$$

Note that reward or punishment doesn't influence all connections. Connections according to the Hebbian theory [Heb49] are only able to learn if a post-synaptic spike followed a pre-synaptic spike in the 50 ms where the movement to the current angle got encoded. Further on only connections from ES to EM neurons are able to learn. (TODO Why find prove). Reward or punishment increases or decreases a weight scale factor of each connection that is able to learn. For more informations look Appendix. TODO just copy picture. A weight lies in the interval [0,5].

Literaturverzeichnis

[CNKL12] George L Chadderdon, Samuel A Neymotin, Cliff C Kerr, and William W Lytton. Reinforcement learning of targeted movement in a spiking neuronal model of motor cortex. *PloS one*, 7(10):e47251, 2012.

[Heb49] Donald Olding Hebb. *The organization of behavior: A neuropsychological approach.* John Wiley & Sons, 1949.

[HH52] Alan L Hodgkin and Andrew F Huxley. A quantitative description of membrane current and its application to conduction and excitation in nerve. *The Journal of physiology*, 117(4):500–544, 1952.

[Izh03] Eugene M Izhikevich. Simple model of spiking neurons. *IEEE Transactions on neural networks*, 14(6):1569–1572, 2003.

[KLM96] Leslie Pack Kaelbling, Michael L Littman, and Andrew W Moore. Reinforcement learning: A survey. *Journal of artificial intelligence research*, 4:237–285, 1996.

[SNR15] Martin Spüler, Sebastian Nagel, and Wolfgang Rosenstiel. A spiking neuronal model learning a motor control task by reinforcement learning and structural synaptic plasticity. In *Neural Networks (IJCNN), 2015 International Joint Conference on*, pages 1–8. IEEE, 2015.