An Introduction to Spiking Neural Networks: Probabilistic Models, Learning Rules, and Applications [Supplementary Material]

Hyeryung Jang*, Osvaldo Simeone*, Brian Gardner†, and André Grüning†

Appendix: Additional Details on the “Examples” section

Batch Learning

For the batch learning example, the training data set is generated by selecting 500 images of digit “1” and “7” from the USPS handwritten digit data set [1]. The test set is similarly obtained by using 125 examples from the USPS data set. We adopt a two-layer SNN, the first layer encoding the input with 256 input neurons and the second the output with 2 output neurons. No hidden neurons exist, thus training is done with the batch SGD rule in (9) for 5000 epochs with constant learning rate $\eta = 0.05$. The model parameters are randomly initialized with uniform distribution in range $[-1, 1]$. We assume the $K_a = K_b = 8$ raised cosine basis functions in Fig. 4 with the maximum filter duration $\tau = T$. As a baseline, a conventional ANN with the same topology and a soft-max output layer is trained for the same epochs with the same learning rate. The average performance is evaluated based on 10 trials with different random seeds.

On-line Learning

In the on-line prediction task, the sequence $\{a_l\}$ is generated from two sequences with duration $T_s = 25$, one from class 1 and the other from class 6, of the SwedishLeaf data set of the UCR archive [2], which are normalized within the range $[0, 1]$. At every $T_s = 25$ time steps, with probability 0.7, an all-zero sequence is selected, and otherwise, one of the two mentioned sequences is selected with equal probability.

In this task, we adopt a fully connected SNN topology that includes $N_{\rm H}$ hidden neurons. In all cases, the SNN contains $K_a = K_b$ weights per synapses, and neurons in which the synaptic and feedback

* H. Jang and O. Simeone are with the Department of Informatics, King's College London, London, United Kingdom (emails: hyeryung.jang@kcl.ac.uk, osvaldo.simeone@kcl.ac.uk).

† B. Gardner and A. Grüning are with the Department of Computer Science, University of Surrey, Guildford, Surrey, United Kingdom (emails: b.gardner@surrey.ac.uk, a.gruning@surrey.ac.uk).

This work was supported in part by the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (grant agreement No. 725731) and by the US National Science Foundation (NSF) under grant ECCS 1710009.

kernels are parameterized using the raised cosine functions in Fig. 4, whose equation can be found in [3]. We use $K_a = K_b = 5$ basis functions with filter duration $\tau = [0.5\Delta T, \Delta T, 3\Delta T, 5\Delta T, 10\Delta T]$. When training, the model parameters are randomly initialized as $\mathcal{N}(0, 0.01)$. We use constant learning rates $\eta = 0.01$ for updating the model parameters; $\kappa = 0.5$ for computing the eligibility traces; and baseline control variates of learning signal with moving average constant 0.01.

In all experiments, the SNN is trained using Algorithm 2 with the addition of a sparsity regularization term. This is obtained by assuming an i.i.d. reference Bernoulli distribution with a desired spiking rate $r \in [0, 1]$, i.e., $\log r(\boldsymbol{h}_{\leq T}) = \sum_{t=0}^{T} \sum_{i \in \mathcal{H}} h_{i,t} \log r + (1 - h_{i,t}) \log(1 - r)$. By adding the regularization term $-\alpha \cdot \mathrm{KL}(q_{\boldsymbol{\theta}^{\mathrm{H}}}(\boldsymbol{h}_{\leq T}|\boldsymbol{x}_{\leq T}) \parallel r(\boldsymbol{h}_{\leq T}))$ to the ELBO in (13), we have the regularized learning signal as

$$\tilde{\ell}(\boldsymbol{x}_{\leq T}, \boldsymbol{h}_{\leq T}) = \sum_{t=0}^{T} \sum_{i \in \mathcal{X}} \log p(x_{i,t}|u_{i,t}) - \alpha \bigg(\sum_{t=0}^{T} \sum_{i \in \mathcal{H}} h_{i,t} \log \frac{\sigma(u_{i,t})}{r} + (1 - h_{i,t}) \log \frac{1 - \sigma(u_{i,t})}{1 - r} \bigg).$$

As a result, the global feedback phase in Algorithm 2 is modified to compute an eligibility trace from the regularized learning signal $\tilde{\ell}$ as

$$\ell_t = \kappa \ell_{t-1} + (1 - \kappa) \bigg(\sum_{i \in \mathcal{X}} \log p(x_{i,t}|u_{i,t}) - \alpha \bigg(\sum_{i \in \mathcal{H}} h_{i,t} \log \frac{\sigma(u_{i,t})}{r} + (1 - h_{i,t}) \log \frac{1 - \sigma(u_{i,t})}{1 - r} \bigg) \bigg).$$

We use a regularization coefficient $\alpha = 1$ and a desired spiking rate $r = 0.1$ for this task.

For time encoding in Fig. 9, we adopt N_{X} number of truncated Gaussian receptive fields [4]. Considering the value a_l smaller than 0.1 as silent signal, Gaussian receptive fields are truncated within range $[0.1, 1]$, which is discretized into N_{X} uniform regions. Each region is assigned to a Gaussian receptive field, in the sense that the receptive field is chosen to have a mean at a center of the corresponding region and a variance of 1.0, and then is normalized within the range $[0, \Delta T]$.

References

[1] J. J. Hull, "A database for handwritten text recognition research," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 16, no. 5, pp. 550–554, 1994.

[2] H. A. Dau, E. Keogh, K. Kamgar, C.-C. M. Yeh, Y. Zhu, S. Gharghabi, C. A. Ratanamahatana, Yanping, B. Hu, N. Begum, A. Bagnall, A. Mueen, and G. Batista, "The ucr time series classification archive," https://www.cs.ucr.edu/ eamonn/time_series_data_2018/, 2018.

[3] J. W. Pillow et al., "Spatio-temporal correlations and visual signalling in a complete neuronal population," *Nature*, vol. 454, no. 7207, p. 995, 2008.

[4] S. M. Bohte, H. La Poutré, and J. N. Kok, "Unsupervised clustering with spiking neurons by sparse temporal coding and multilayer rbf networks," *IEEE Transactions on Neural Networks*, vol. 13, no. 2, pp. 426–435, 2002.