Improving eligibility propagation using Izhikevich neurons in a multilayer RSNN.

Presentation 5: Evaluation meeting 1

Werner van der Veen
(w.k.van.der.veen.2@student.rug.nl)

December 7, 2020

A quick project overview

Task: using eligibility propagation to classify phonemes.

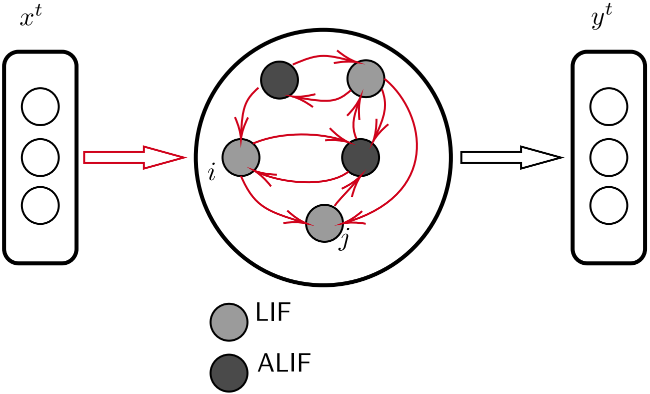
Challenges:

- Desired bio-plasubility requires online and local learning rules.
- Spiking and reccurent network: no good learning algorithm yet.

My own contribution: investigate the benefits of

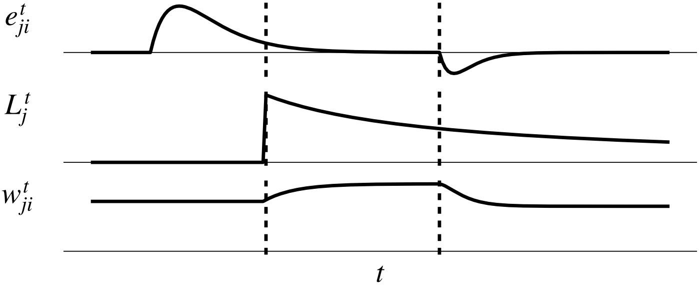
- stacking multiple recurrent layers;
- using Izhikevich neurons;
- various regularization methods.

Original network architecture



The e-prop learning algorithm – Main equation

$$\frac{dE}{dW_{ji}} = \sum_t L_j^t \cdot e_{ji}^t$$



The e-prop learning algorithm – proof (L_j^t)

$$\frac{dE}{dW_{ji}} = \sum_t L_j^t \cdot e_{ji}^t$$

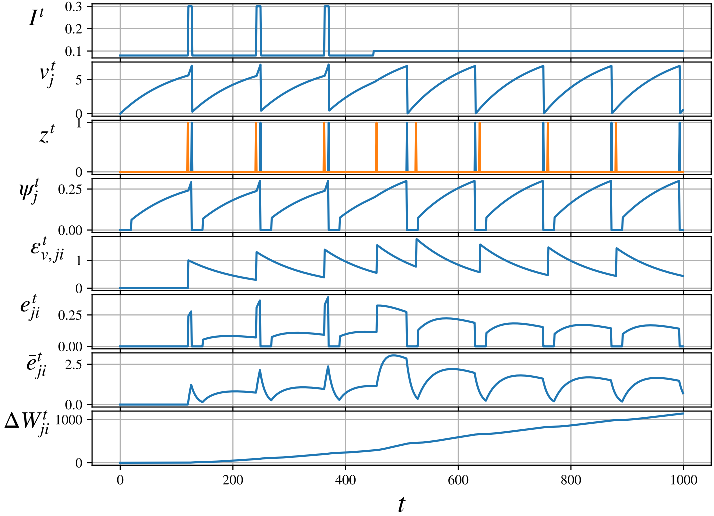
$$\begin{aligned} L_j^t &\stackrel{\text{def}}{=} \frac{\partial E}{\partial z_j^t} \\ &= \sum_k B_{jk} \left(\pi_k^t - \pi_k^{*,t} \right) \end{aligned}$$

The e-prop learning algorithm – proof (e_{ji}^t)

$$e_{ji}^t \stackrel{\text{def}}{=} \frac{\partial z_j^t}{\partial \mathbf{h}_j^t} \underbrace{\sum_{t' \leq t} \frac{\partial \mathbf{h}_j^t}{\partial \mathbf{h}_j^{t-1}} \cdots \frac{\partial \mathbf{h}_j^{t'+1}}{\partial \mathbf{h}_j^{t'}} \cdot \frac{\partial \mathbf{h}_j^{t'}}{\partial W_{ji}}}_{\stackrel{\text{def}}{=} \boldsymbol{\varepsilon}_{ji}^t}$$

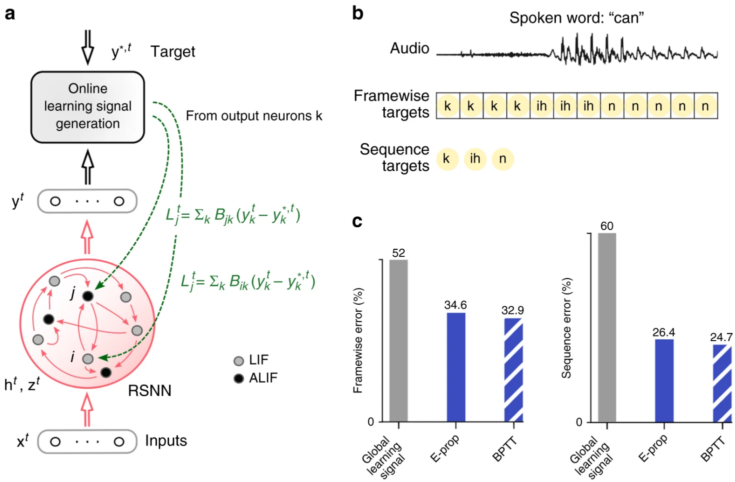
$$= \psi_j^t \bar{z}_i^{t-1} \qquad \text{(for LIF neurons in Bellec et al)}$$

The e-prop learning algorithm – Original LIF



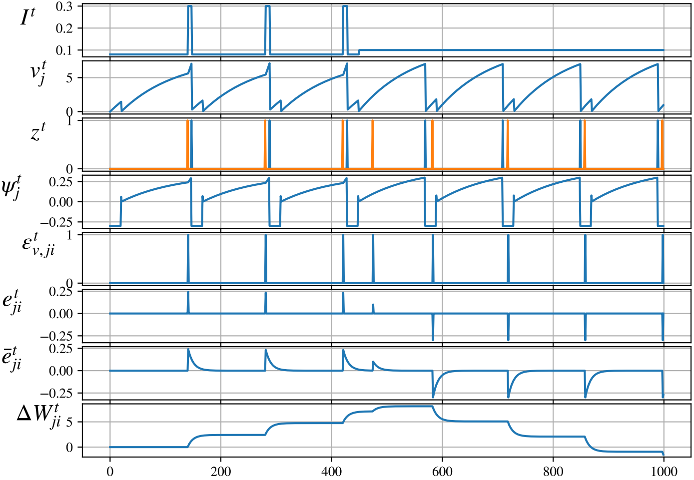
Previous results

Bellec, G., Scherr, F., Subramoney, A., Hajek, E., Salaj, D., Legenstein, R., & Maass, W. (2020). A solution to the learning dilemma for recurrent networks of spiking neurons.



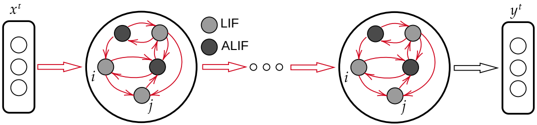
The e-prop learning algorithm – STDP-corrected LIF

Traub, M., Butz, M. V., Baayen, R. H., & Otte, S. (2020, September). Learning Precise Spike Timings with Eligibility Traces. In International Conference on Artificial Neural Networks (pp. 659-669). Springer, Cham.



My proposed contribution

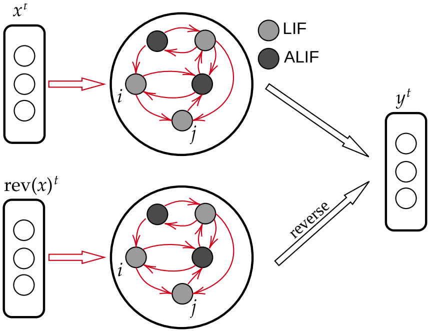
– Implement the algorithm in a *multi-layer* SRNN.



– Evaluate model designs such as neuron type (e.g. Izhikevich); synaptic delay; and regularization such as metaplasticity and synaptic scaling.

Work done so far

- ☑ Implement Bellec's model with
  - ☑ 1-layer e-prop RSNN with ALIF neurons;
  - ☑ TIMIT conversion to 13 MFCCs and their first and second derivatives;
  - ☑ L^2 and firing rate regularization;
  - ☑ Adam optimizer;
  - ☑ Bidirectional network.
- ☐ Obtain Bellec's performance;
- ☐ Analyze effects of
  - ☑ N-layered network;
  - ☐ Izhikevich neurons;
  - ☐ Metaplasticity;
  - ☑ Synaptic scaling;
  - ☑ Traub's fix;
  - ☐ Synaptic delays.



Work since last meeting

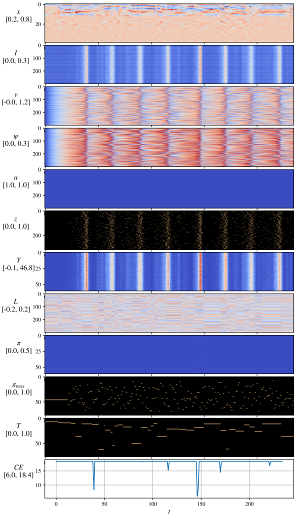
- ☑ Implemented bidirectional network;
- ☑ Exponentially improved space complexity by optionally not tracking synapse variables over time or epochs;
- ☑ Implemented Bayesian hyperparameter optimization using Gaussian processes;

Current observations

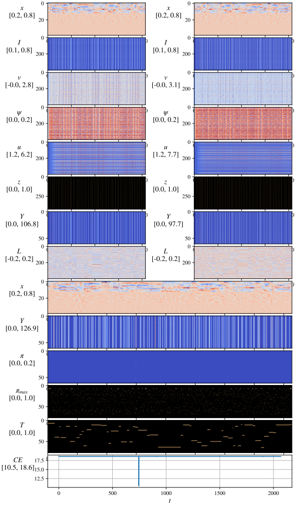
-
$$\lim_{e \to \infty} W^e_{\text{out}} = \mathbf{0}$$
$$\lim_{e \to \infty} b^e_{\text{out}} = \mathbf{0}$$

- Spike frequency normal but heavily affected by weight initialization. Weights in uniform range of $[0, \frac{C}{N}]$ work well, where N is number of afferent connections.

Typical results (Undesired pulsation)



Typical results (Bellec initialization)



Current configuration toolbox

- E-prop type (random, symmetric, adaptive) (for updating B_j^t)
- Optimizer (Adam, SGD)
- Traub fix
- Fraction ALIF
- Uni- or bidirectional
- Target delay
- Various e-prop model constants
- Initial weight scaling
- Weight decay
- L^2 regularization
- Firing rate regularization (strength & target Hz)
- Weight dropout rate
- Adam optimizer constants
- Number of layers
- Nodes per layer
- Input repeats
- Batch size

Hyperparameter sweep results

(show results)

My questions

- Thesis template? (`.tex`)
- Any questions or suggestions?