Supplemental Material for Spiking Reinforcement Learning Enhanced by Bioinspired Event Source of Multi-dendrite Spiking Neuron and Dynamic Thresholds

Xingyue Liang, Qiaoyun Wu, *Member, IEEE*, Yun Zhou, Chunyu Tan, *Member, IEEE*, Hongfu Yin and Changyin Sun, *Senior Member, IEEE*

I. Appendices

The distal dendrite receives spikes represented by fractions, obtained through population encoding based on quantile fractions of the state spikes. To represent fractional value with population neural spikes, the firing rate of the j-th neuron for the value ρ_k is

$$r_{kj} = \Gamma_j e^{\frac{-(\rho_k-\lambda_j)^2}{2\sigma_j^2}}, \tag{1}$$

where ρ_k is k-th fractional value, each neuron in the population has a Gaussian receptive field (λ_j, σ_j) with two parameters of mean and standard deviation. The spiking activity of the neurons is a Poisson process. The quantile fraction ρ_k is encoded in the population spike

$$P_{\rho_k} = \left\langle P_{kj} \sim p_{kj} = \frac{(r_{kj}\Delta t)^2 e^{-r_{kj}\Delta t}}{n!}, j = 1, 2, ..., m \right\rangle, \tag{2}$$

where p_{kj} is neuron j firing n spikes in time interval Δt with probability.

A. Loss Function

SCNN and final fully connected SNN weights were optimized by minimizing Huber's loss function.

$$HL(\rho, \delta_{ij}) = |\rho - (1 - H(\delta_{ij}))| \frac{L(\eta, \delta_{ij})}{\eta}, \tag{3}$$

$$L(\eta, \delta_{ij}) = \begin{cases} \frac{1}{2}\delta_{ij}^2, & \text{if} |\delta_{ij}| \le \eta \\ \eta(|\delta_{ij}| - \frac{1}{2}\eta), & \text{otherwise.} \end{cases} \tag{4}$$

$$H(\delta_{ij}) = \begin{cases} 0, & \text{if } \delta_{ij} \le 0 \\ 1, & \text{otherwise.} \end{cases} \tag{5}$$

Where $\delta_{ij}^t = r_t + F^{(-1)}(\hat{\rho}_i \mid s_{t+1}, a_{t+1}) - F^{(-1)}(\hat{\rho}_j \mid s_t, a_t)$ is the temporal difference (TD) error for quantile value distribution.

The weight parameter W is optimized through the network as a whole, by minimizing the Wasserstein loss of the quantile value distribution:

$$WL(\rho) = \sum_{k=0}^{N-1} \int_{\rho_k}^{\rho_{k+1}} |F^{-1}(\theta) - F^{-1}(\hat{\rho}_k)| d\theta, \tag{6}$$

$$\frac{\partial WL(\rho)}{\partial \rho_k} = 2F^{-1}(\rho_k) - F^{-1}(\hat{\rho}_k) - F^{-1}(\hat{\rho}_{k-1}). \tag{7}$$

We obtain the derivative of $WL(\rho)$ with respect to W_k^f,

$$\begin{aligned} \frac{\partial WL(\rho)}{\partial W_k^f} &= \sum_{n=1}^{N} \frac{\partial WL(\rho)}{\partial \rho_n} \frac{\partial \rho_n}{\partial \Gamma_k} \frac{\partial \Gamma_k}{\partial W_k^f} \\ &= \frac{1}{T} \sum_{n=1}^{N} \sum_{h=0}^{n-1} \sum_{t=0}^{T} \frac{\partial WL(\rho)}{\partial \rho_n} \Delta_{k,h} O_t^s, \end{aligned} \tag{8}$$

where $\Delta_{k,h} = (-\xi_h \xi_k) + (N - k + 1)\xi_k(1 - \xi_k)$.

The dendritic synaptic weights are trained with backpropagation from the fully connected SNN, as follows:

$$\begin{aligned} \frac{\partial HL(\rho, \delta_{ij})}{\partial \omega^\alpha} &= \frac{\partial HL(\rho, \delta_{ij})}{\partial o_t} \frac{\partial o_t}{\partial V(t)} \frac{\partial V(t)}{\partial u_{s,t}^\alpha} \frac{\partial u_{s,t}^\alpha}{\partial \omega^\alpha} \\ &= \sum_{t=1}^{T} \sum_{h=0}^{t-1} \varphi \frac{2\tau}{4 + (\pi\tau V(t))^2} \Lambda^{T-t} \\ &\quad \cdot \frac{g_\alpha}{g\tau} (1 - \frac{1}{\tau_\alpha})^{t-h-1} \frac{1}{\tau_\alpha} O_h^s \\ &= \sum_{t=1}^{T} \sum_{h=0}^{t-1} \frac{2 g_\alpha \varphi}{g \tau_\alpha (4 + (\pi\tau V(t))^2)} \Lambda^{T-t} \\ &\quad \cdot (1 - \frac{1}{\tau_\alpha})^{t-h-1} O_h^s; \end{aligned} \tag{9}$$

$$\begin{aligned} \frac{\partial HL(\rho, \delta_{ij})}{\partial \omega^\beta} &= \frac{\partial HL(\rho, \delta_{ij})}{\partial o_t} \frac{\partial o_t}{\partial V(t)} \frac{\partial V(t)}{\partial u_{s,t}^\beta} \frac{\partial u_{s,t}^\beta}{\partial \omega^\beta} \\ &= \sum_{t=1}^{T} \sum_{h=0}^{t-1} \varphi \frac{2\tau}{4 + (\pi\tau V(t))^2} \Lambda^{T-t} \\ &\quad \cdot \frac{g_\beta}{g\tau} (1 - \frac{1}{\tau_\beta})^{t-h-1} \frac{1}{\tau_\beta} P_{\rho,h} \\ &= \sum_{t=1}^{T} \sum_{h=0}^{t-1} \frac{2 g_\beta \varphi}{g \tau_\beta (4 + (\pi\tau V(t))^2)} \Lambda^{T-t} \\ &\quad \cdot (1 - \frac{1}{\tau_\beta})^{t-h-1} P_{\rho,h}; \end{aligned} \tag{10}$$

$$
\begin{aligned}
\frac{\partial HL(\rho,\delta_{ij})}{\partial \omega^{\gamma}} &= \frac{\partial HL(\rho,\delta_{ij})}{\partial o_t}\frac{\partial o_t}{\partial V(t)}\frac{\partial V(t)}{\partial u_{s,t}^{\gamma}}\frac{\partial u_{s,t}^{\gamma}}{\partial \omega^{\gamma}} \\
&= \sum_{t=1}^{T}\sum_{h=0}^{t-1}\varphi\frac{2\tau}{4+(\pi\tau V(t))^2}\Lambda^{T-t} \\
&\quad \cdot \frac{g_{\gamma}}{g\tau}(1-\frac{1}{\tau_{\gamma}})^{t-h-1}\frac{1}{\tau_{\gamma}}M_h \qquad (11) \\
&= \sum_{t=1}^{T}\sum_{h=0}^{t-1}\frac{2g_{\gamma}\varphi}{g\tau_{\gamma}(4+(\pi\tau V(t))^2)}\Lambda^{T-t} \\
&\quad \cdot (1-\frac{1}{\tau_{\gamma}})^{t-h-1}M_h;
\end{aligned}
$$

where $\varphi = \frac{\partial HL(\rho,\delta_{ij})}{\partial o_t}$, and $\Lambda = (1-\frac{1}{\tau}-\frac{g_{\alpha}+g_{\beta}+g_{\gamma}}{g\tau})^{T-t}$.