2022年 江苏省人工智能学术会议

Learning And Mining from DatA

Bifurcation Spiking Neural Network

Journal: Journal of Machine Learning Research (2021)

Authors: **Shao-Qun Zhang**, Zhao-Yu Zhang, and Zhi-Hua Zhou

Speaker: **Shao-Qun Zhang (张绍群)**

Institution: Nanjing University

Date: 2022/11/19

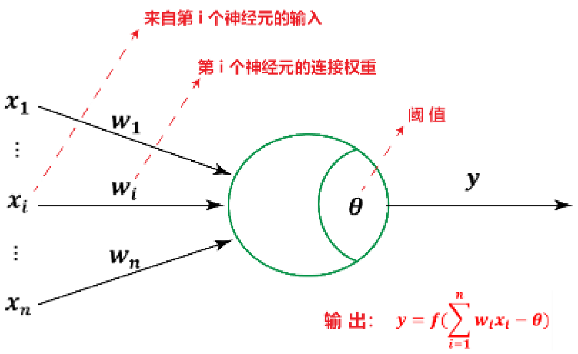
Neural networks learning

Neural Network Learning = Neural Network Model + Learning Algorithm

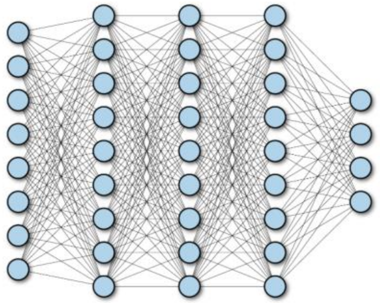
Neuron Model + Network Architecture

e.g., BP algorithm

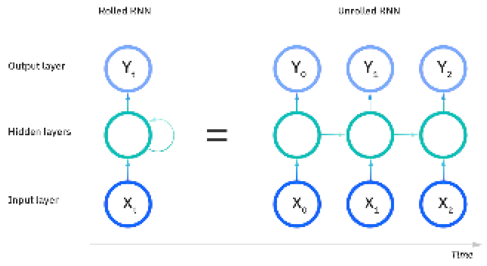
MP Neuron Model, 1943



e.g., feed-forward



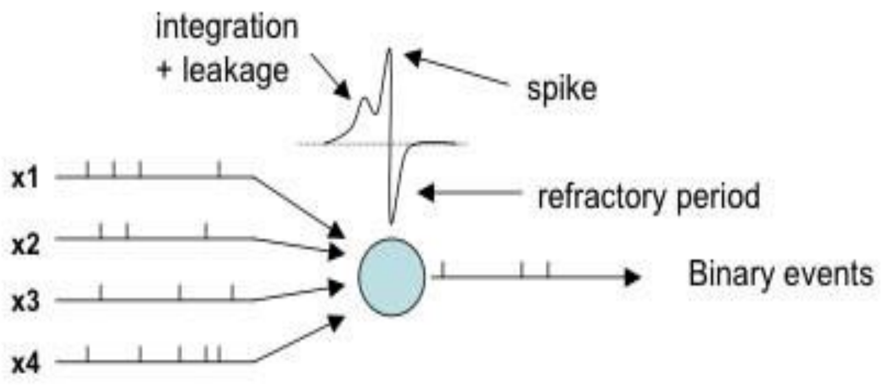
e.g., recurrent



little explored

extensively studied

Spiking neuron model

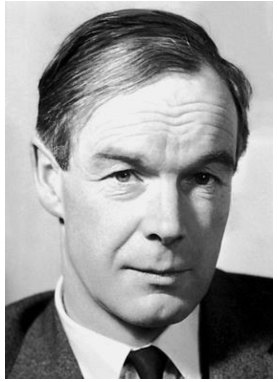


spiking neuron model

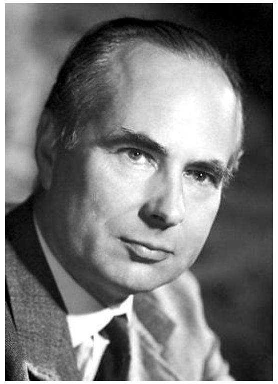
Leaky Integrate-and-Fire equation (LIF)

$$\tau_m \frac{\mathrm{d}u}{\mathrm{d}t} = -u + \sum_{i=1}^{m} w_i x_i$$

denotes the "leak" term



Alan L. Hodgkin
(1914 – 1998)
1963 Nobel Prize



Andrew F. Huxley
(1917 – 2012)
1963 Nobel Prize

Louis Lapicque
(1866 – 1952)

Marcelle Lapicque
(1873 – 1960)

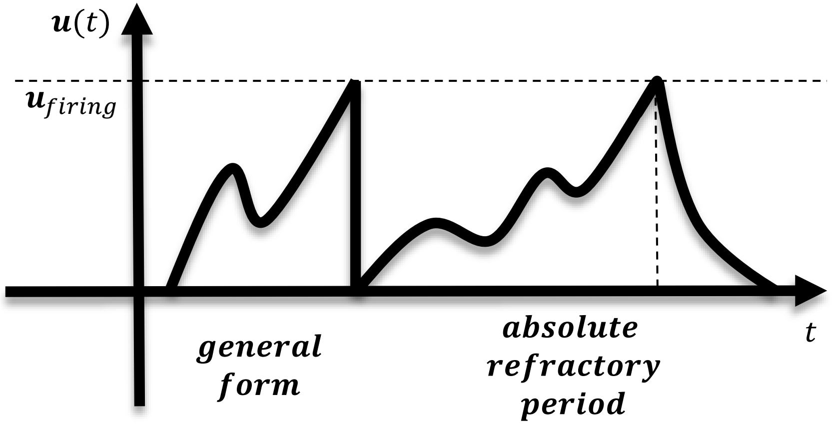
Computation Procedure

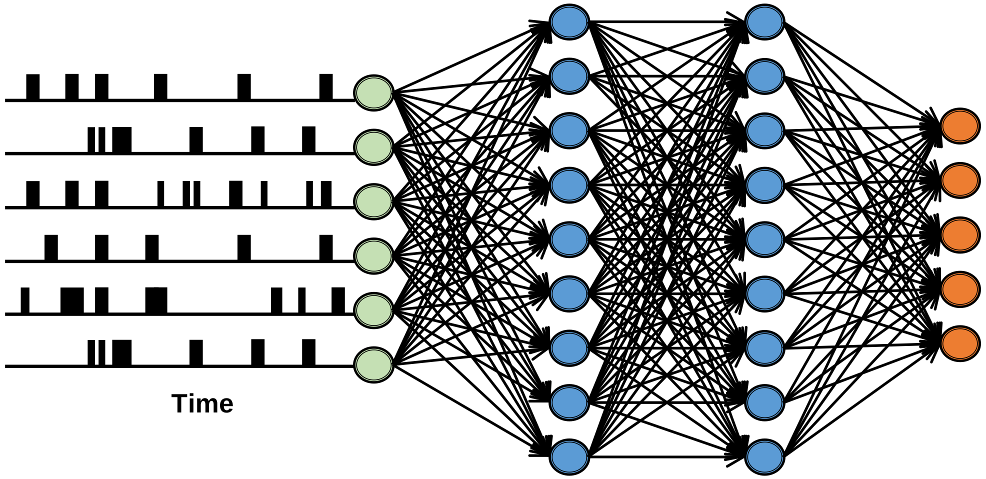
Step 1: Integration

$$\tau_m \frac{\mathrm{d}u}{\mathrm{d}t} = -u + \sum_{i=1}^{m} w_i x_i$$

Step 2: Firing

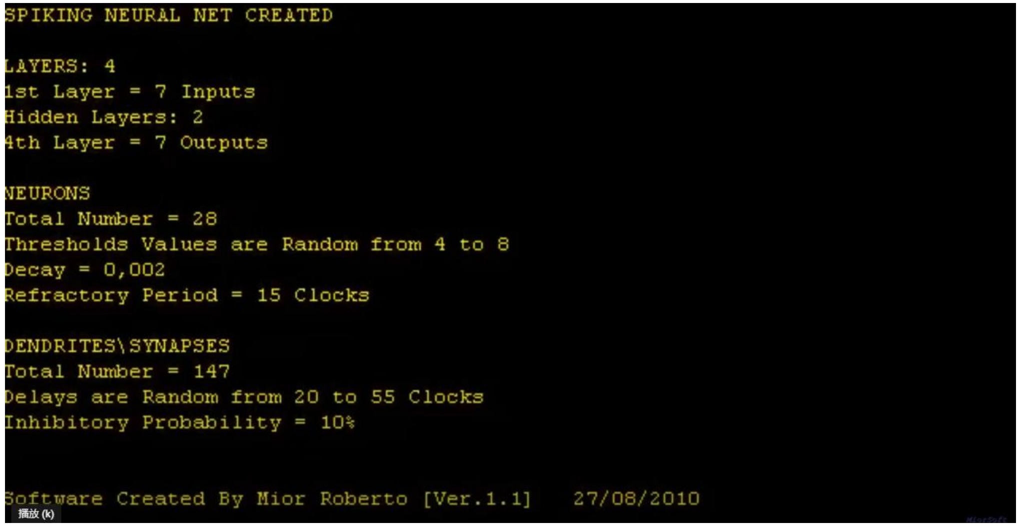
$$f_e : u \to S, \quad \text{where} \quad S(t) \triangleq \left\lfloor \frac{u(t)}{u_{firing}} \right\rfloor$$





spike sequence hidden output

Raster Illustration



SPIKING NEURAL NET CREATED

LAYERS: 4
1st Layer = 7 Inputs
Hidden Layers: 2
4th Layer = 7 Outputs

NEURONS
Total Number = 28
Thresholds Values are Random from 4 to 8
Decay = 0,002
Refractory Period = 15 Clocks

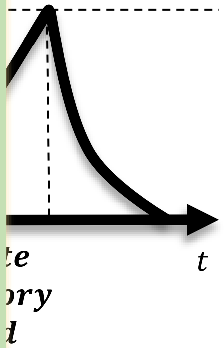
DENDRITES\SYNAPSES
Total Number = 147
Delays are Random from 20 to 55 Clocks
Inhibitory Probability = 10%

Software Created By Mior Roberto [Ver.1.1] 27/08/2010

Computation Pr

Step 1: Integration

$$\tau_m \frac{\mathrm{d}u}{\mathrm{d}t} = -$$

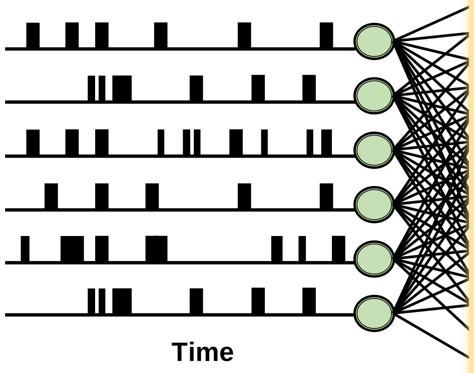


Potentials:

- Sparse computation
- Event-driven
- Temporal representation

Step 2: Firing

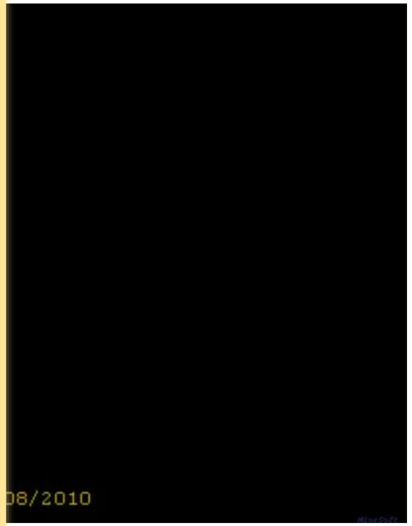
$f_e : u \to S,$ where ... $\lfloor u_{firing} \rfloor$



spike sequence

Two key issues:

- Theoretical Understandings
- Killer Applications



Investigation from dynamical systems

$$\tau_m \frac{\mathrm{d}u}{\mathrm{d}t} = -u + \sum_{i=1}^{m} w_i x_i$$

multiply u on both sides

$$\frac{\tau_m}{2} \frac{\mathrm{d}u^2}{\mathrm{d}t} = -u^2 + \left(\sum_{i=1}^{m} w_i x_i \right) u$$

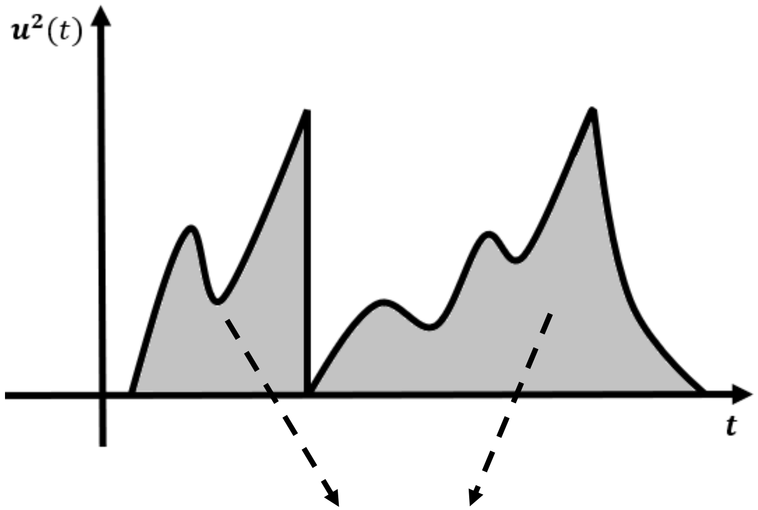
$$-\frac{2}{\tau_m} u^2 = \frac{\mathrm{d}u^2}{\mathrm{d}t} - \frac{2}{\tau_m} \left(\sum_{i=1}^{m} w_i x_i \right) u$$

energy estimation

theorem

energy function

$$\mathcal{H}(u,t) \stackrel{\text{def}}{=} u^2 - \frac{2}{\tau_m} \int_t \left(\sum_{i=1}^{m} w_i x_i(t) \right) u(t) \, \mathrm{d}t$$



$\mathcal{H}(u,t)$ denotes the integral of motion u^2.

Bifurcation

energy derivative

$$\frac{\mathrm{d}\mathcal{H}}{\mathrm{d}t} = 2u\left[\frac{\mathrm{d}u}{\mathrm{d}t} + \frac{1}{\tau_m}\left(\sum_{i=1}^{m} w_i x_i\right)\right] = -\frac{2}{\tau_m}u^2$$

$$\begin{cases} \frac{1}{\tau_m} < 0 & \frac{\mathrm{d}\mathcal{H}}{\mathrm{d}t} > 0 & \text{energy increases (扩散系统)} \\ \frac{1}{\tau_m} = 0 & \frac{\mathrm{d}\mathcal{H}}{\mathrm{d}t} = 0 & \text{energy unchanged (保守系统)} \\ \frac{1}{\tau_m} > 0 & \frac{\mathrm{d}\mathcal{H}}{\mathrm{d}t} < 0 & \text{energy decreases (耗散系统)} \end{cases}$$

relies on τ_m,
not on connection weight w_i

Theorem 1. Given the initial condition $u_0 = 0$, the dynamical system led by one layer of LIF neurons is a bifurcation dynamical system, and τ_m is the corresponding bifurcation hyper-parameter.

An improper setting of τ_m will disable the possibility of proper learning.

Before learning, one can hardly know how to set a proper τ_m.

To make the bifurcation hyper-parameter adaptive

- Gradient-based approach
  - Alternating gradient optimization

    lack of convergence guarantee

  - Pack τ_m and w_i as one parameter

    easy to fall into the problems of gradient explosion and vanishing

LIF solution: $u(t) = \sum_{i=1}^{m} w_i \left[\int_{t'}^{t} \exp\left(\frac{t'-s}{\tau_m}\right) x_i(s)\, \mathrm{d}s \right]$

entangled

Gradients:

$$\begin{cases} \nabla \tau_m \propto \sum_{i=1}^{m} w_i \left[\int_{t'}^{t} \exp\left(\frac{t'-s}{\tau_m}\right) (s-t') x_i(s)\, \mathrm{d}s \right], \\ \nabla w_i \propto \left[\int_{t'}^{t} \exp\left(\frac{t'-s}{\tau_m}\right) x_i(s)\, \mathrm{d}s \right] \end{cases}$$

integration on the time interval

- Zero-order approach

  succeeds on an apposite initialization and larger computation and storage

Our solution: Introducing the mutual promotion

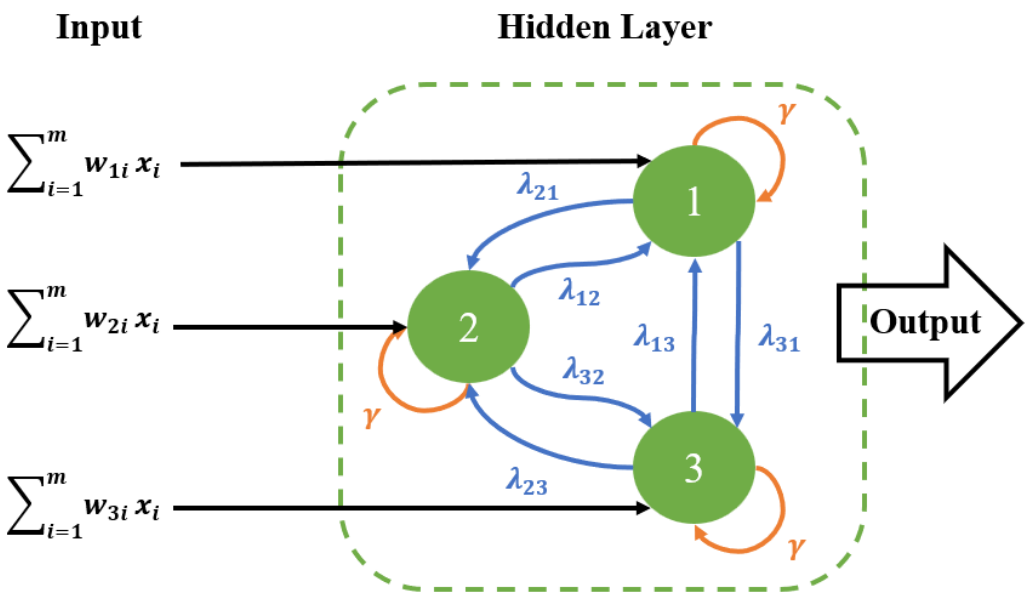
BSNN (Bifurcation Spiking Neural Network)

- employs the self-connection structure

$$\frac{\partial u_k}{\partial t} = -\frac{1}{\tau_m} u_k + \sum_{j \neq k} \lambda_{kj} u_j + \frac{1}{\tau_m} \sum_{i=1}^{m} w_{ki} x_i$$

mutual promotion
between neurons

- leaves learnable connection weights and mutual promotion parameters



BSNN (Bifurcation Spiking Neural Network)

energy derivative

$$\frac{\mathrm{d}\mathcal{H}_{\mathrm{B}}}{\mathrm{d}t} = 2\begin{pmatrix} u_1 \\ u_2 \\ \vdots \\ u_n \end{pmatrix}^{\top} \begin{pmatrix} -\frac{1}{\tau_m} & \lambda_{12} & \cdots & \lambda_{1n} \\ \lambda_{21} & -\frac{1}{\tau_m} & \cdots & \lambda_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ \lambda_{n1} & \lambda_{n2} & \cdots & -\frac{1}{\tau_m} \end{pmatrix} \begin{pmatrix} u_1 \\ u_2 \\ \vdots \\ u_n \end{pmatrix}$$

hyper-parameter τ_m + learnable matrix $\boldsymbol{\lambda}$

upper bound: 2^{n-1}

lower bound: $\Omega((n+1)^2 \ln(n+1))$

enough solutions

$$\frac{\mathrm{d}\mathcal{H}_{\mathrm{B}}}{\mathrm{d}t} >, \ =, \ \text{or} \ < 0 \ ?$$

allow gradient calculation

SRM-based training algorithms

Theorem 2. If the bifurcation parameters λ_{ij} are all great than 0, there are at most 2^{n-1} bifurcation solutions, where n is the number of hidden spiking neurons.

BSNN is calculable and has enough solutions to be adaptive to data/environment.

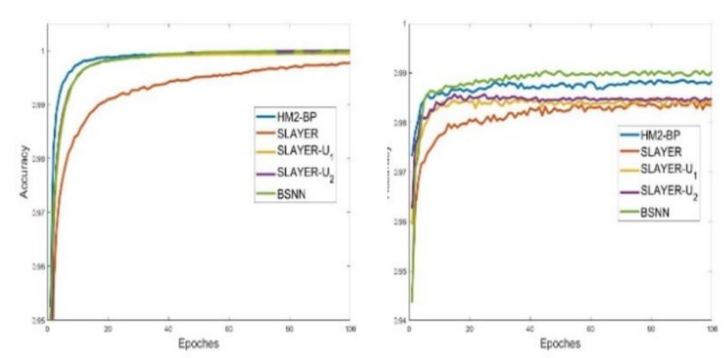
Experimental Results

Best Accuracy

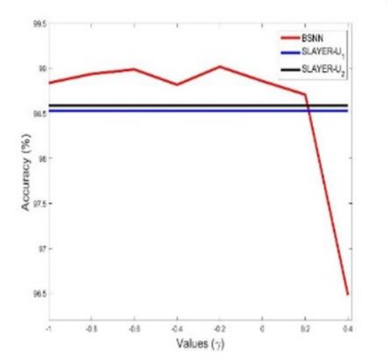
Table 2: The comparative performance of the contenders and BSNN.

Data Sets	Contenders	Accuracy (%)	Setting	Control Rate (γ)	Epochs
MNIST	Deep SNN (O'Connor and Welling, 2016)	97.80	28×28-300-300-10 ♠	-	50
	Deep SNN-BP (Lee et al., 2016)	98.71	28×28-800-10	-	200
	SNN-EP ♡	97.63	28×28-500-10	-	25
	HM2-BP (Jin et al., 2018)	98.84 ± 0.02	28×28-800-10	-	100
	SNN-L (Rezaabad and Vishwanath, 2020)	98.23 ± 0.07	28×28-1000-R28-10	-	-
	SLAYER (Shrestha and Orchard, 2018)	98.39 ± 0.04	28×28-500-500-10	-	50
	SLAYER-U_1 ♣	98.53 ± 0.03	28×28-500-500-10	-	-
	SLAYER-U_2	98.59 ± 0.01	28×28-500-500-10	-	-
	BSNN (this work)	**99.02 ± 0.04**	28×28-500-500-10	-0.21	50
N-MNIST	SKIM (Cohen et al., 2016)	92.87	2*28×28-10000-10	-	-
	Deep SNN-BP	98.78	2*28×28-800-10	-	200
	HM2-BP	98.84 ± 0.02	2*28×28-800-10	-	60
	SLAYER	98.89 ± 0.06	2*28×28-500-500-10	-	50
	SLAYER-U_1	99.01 ± 0.01	2*28×28-500-500-10	-	-
	SLAYER-U_2	99.07 ± 0.02	2*28×28-500-500-10	-	-
	BSNN (this work)	**99.24 ± 0.12**	2*28×28-500-500-10	-0.49	50
Fashion-MNIST	HM2-BP	88.99	28×28-400-400-10	-	15
	SLAYER	88.61 ± 0.17	28×28-500-500-10	-	50
	SLAYER-U_1	90.53 ± 0.04	28×28-500-500-10	-	-
	SLAYER-U_2	90.61 ± 0.02	28×28-500-500-10	-	-
	ST-RSBP (Zhang and Li, 2019)	90.00 ± 0.13	28×28-400-R400-10 ◇	-	30
	BSNN (this work)	**91.22 ± 0.06**	28×28-500-500-10	-0.32	50
EMNIST	eRBP (Neftci et al., 2017)	78.17	28×28-200-200-47	-	30
	HM2-BP	84.43 ± 0.10	28×28-400-400-10	-	20
	SNN-L	83.75 ± 0.15	28×28-1000-R28-10	-	-
	SLAYER	85.73 ± 0.16	28×28-500-500-47	-	50
	SLAYER-U_2	86.62 ± 0.03	28×28-500-500-47	-	50
	BSNN (this work)	**87.51 ± 0.23**	28×28-500-500-47	-0.37	50

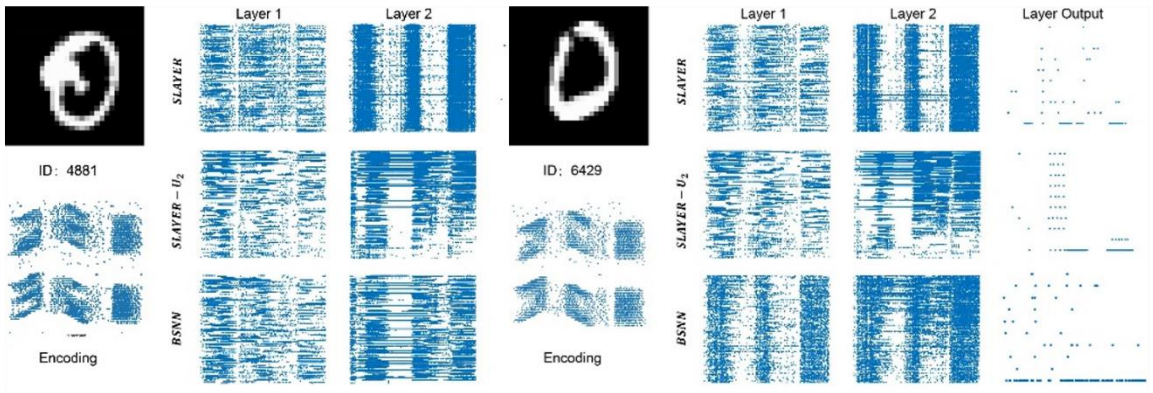
Fastest Convergency



Robustness



spike raster plots



(a) example 4881

(b) example 6429

在基准任务上表现优越！

Further Research

adaptive dynamical systems

Theorem 1. *If the bifurcation hyper-parameters λ_{ij} are all great than 0, there are at most 2^{n-1} bifurcation solutions in BS the neuron model.*

Theorem 2. *Let $H(n)$ denote the maximum possible number of bifurcation solutions of dynamical systems in BS the neuron model with n-order polynomial bifurcation fields in BS the neuron model. Then $H(n)$ is calculable, and we have $H(n) \in \Omega((n+1)^2 \ln(n+1))$.*

calculable
enough adaptive solutions!

- **Shao-Qun Zhang** and Zhi-Hua Zhou. Theoretically Provable Spiking Neural Networks. In: Advances in Neural Information Processing Systems 36 **(NeurIPS'22),** in press. 2022.
- Gao Zhang and **Shao-Qun Zhang**. Structural Stability of Spiking Neural Networks. arXiv:2207.04876. 2022.

approximation to discrete-time dynamical systems

Theorem 3. *Let $K \subset \mathbb{R}^m$ be a compact set. If the spike excitation function f_e is l-finite and $w_i \in \mathbb{R}$ for any $i \in [n]$, then for all $r \in [l]$, there exists some time t such that the set of IFR functions $f(\cdot, t) : K \to \mathbb{R}$ of the form $f(x, t) = \sum_{i \in [n]} w_i f_i(x, t)$, is dense in $C^r(K, \mathbb{R})$.*

universal approximation

Theorem 4. *Given a compact set $K^m \subset \mathbb{R}^m$, a probability measure μ, and a radial function $g : K^m \to \mathbb{R}$. For some apposite spike excitation function f_e and any $\epsilon > 0$, there exists some time t such that the radial function g can be well approximated by a one-hidden-layer scSNN of $\mathcal{O}(Cm^{15/4})$ spiking neurons, that is,*

$$\| f(x, t) - g(x) \|_{L^2(\mu)} < \epsilon \ .$$

polynomial parameter complexity

Theorem 5. *Let $n \geq m$. Let $x_k(t) \in \pi(\lambda_k)$ and $\mathbb{E}(x_k) = \lambda_k$ for $\lambda_k > 0$, $k \in [m]$, and $t \in [T]$. If $\mathbf{V}$ is a non-degenerate matrix, then for any $\epsilon > 0$ and matrix $\mathbf{G} \in \mathbb{R}^{m \times n}$ with $\|\mathbf{G}\|_2 < \infty$, when the time complexity satisfies*

$$T \geq \Omega\left(\frac{\sqrt{n}\ \|\mathbf{G}\|_2}{\epsilon \sqrt{\|\mathbf{V}\|_2}} \right) ,$$

there exists some one-hidden-layer scSNNs with the IFR vector $f = (f_1, \ldots, f_n)^\top$ such that

$$\| \mathbb{E}_x \left[\mathbf{V} f(x, T) \right] - \mathbf{G}\lambda \|_2 < \epsilon \ .$$

polynomial time complexity

上述结论首次证明了脉冲神经元模型的数理功能和计算优势！

For details

Focus:

- **Shao-Qun Zhang**, Zhao-Yu Zhang, and Zhi-Hua Zhou. Bifurcation Spiking Neural Network. Journal of Machine Learning Research **(JMLR),** 22(253):1-21. 2021.

Others:

- **Shao-Qun Zhang** and Zhi-Hua Zhou. Theoretically Provable Spiking Neural Networks. In: Advances in Neural Information Processing Systems 35 **(NeurIPS'22),** in press. 2022.
- Gao Zhang and **Shao-Qun Zhang**. Structural Stability of Spiking Neural Networks. arXiv:2207.04876. 2022.

You can find the detailed proofs and codes on the Homepage of Shao-Qun Zhang.

Contact me by email at zhangsq@lamda.nju.edu.cn

Thank you !

Q & A