Flexible Transmitter Network (NCJ'21)

Shao-Qun Zhang and Zhi-Hua Zhou

LAMDA Group, Nanjing University, Nanjing, China

Contact

zhangsq@lamda.nju.edu.cn
zhouzh@lamda.nju.edu.cn

Biological Neuron Structure

Neural Network Learning = Neural Network Model + Learning Algorithm

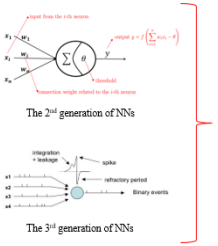
e.g., BP algorithm

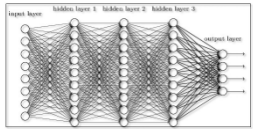
Neuron Model + Network Architecture

e.g., MP model

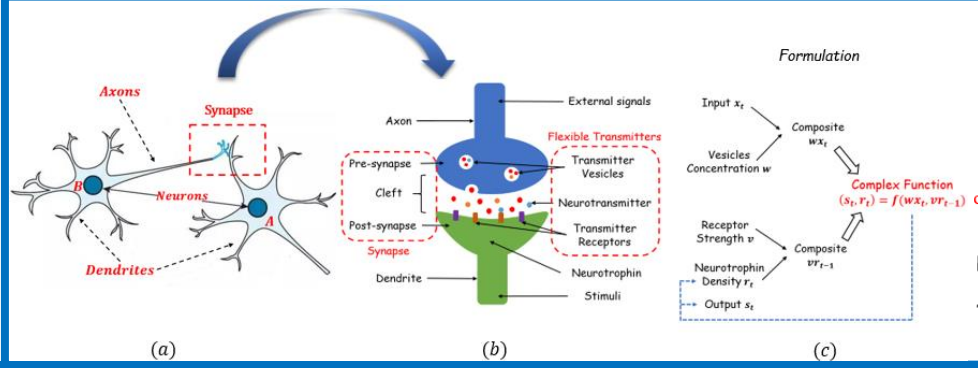
e.g., feed-forward

Spiking Neural Model





Flexible Transmitter Network



A two-variable two-valued function:

memory variables, implicit → Local Recurrent Updating

$$(s_t, r_t) = f(w\,x_t, v\,r_{t-1})$$

output stimuli, explicit — input signals — parameters

w corresponds to x_t

v corresponds to r_{t-1}

Requirements:

s_t and r_t are functions of both x_t and r_{t-1}, respectively.

Contributions

M-P Model:

megascopic view for modeling the neuron

Spiking Model:

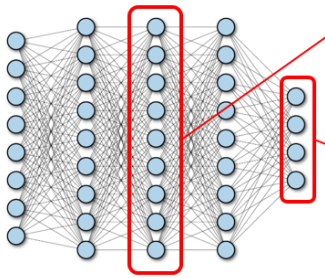
microcosmic view for modeling the synaptic plasticity

Our Work:

provides a meso-corporatism view

Complex Reaction

Full-connected Feed-forward Architecture:



Layer-vectorized: → implicit variables, unsupervised

$$S_t^l + M_t^l\,\boldsymbol{i} = \sigma[\,(\boldsymbol{W}^l S_t^{l-1} - \boldsymbol{V}^l R_{t-1}^l) + (\boldsymbol{W}^l S_t^{l-1} + \boldsymbol{V}^l R_{t-1}^l)\boldsymbol{i}\,]$$

explicit variables, supervised

Time-dependent Cost Function:

$$E(\boldsymbol{W}, \boldsymbol{V}) = \frac{1}{2}\int_{t=1}^{T}\sum_{i=1}^{n_L}(\widehat{y}_t^i - y_t^i)^2\,dt$$

Gradient Calculation:

$$\left(\nabla_{\boldsymbol{W}^l}E, \nabla_{\boldsymbol{V}^l}E\right) = \int_{t=1}^{T}\left(\nabla\boldsymbol{W}_t^l, \nabla\boldsymbol{V}_t^l\right)\,dt$$

What's the effect of m_t in training ?

$$\left(\nabla_{\boldsymbol{W}^l}E, \nabla_{\boldsymbol{V}^l}E\right) = \int_{t=1}^{T}\left(\nabla\boldsymbol{W}_t^l, \nabla\boldsymbol{V}_t^l\right)\,dt$$

Layer-dependent

Time-dependent

For $S_t^l + R_t^l\,\boldsymbol{i} = \sigma[\,(\boldsymbol{W}^l S_t^{l-1} - \boldsymbol{V}^l R_{t-1}^l) + (\boldsymbol{W}^l S_t^{l-1} + \boldsymbol{V}^l R_{t-1}^l)\boldsymbol{i}\,]$

$$\nabla\boldsymbol{W}_t^l \sim \frac{\partial S_t^l}{\partial \boldsymbol{W}^l} \sim \frac{\partial \boldsymbol{W}^l S_t^{l-1}}{\partial \boldsymbol{W}^l} - \frac{\partial R_{t-1}^l}{\partial \boldsymbol{W}^l}$$

$$\nabla\boldsymbol{V}_t^l \sim \frac{\partial S_t^l}{\partial \boldsymbol{V}^l} \sim 0 + \frac{\partial \boldsymbol{V}^l R_{t-1}^l}{\partial \boldsymbol{V}^l}$$

Layer-dependent Time-dependent

where,

$$\frac{\partial R_{t-1}^l}{\partial \boldsymbol{W}^l} \sim \frac{\partial \boldsymbol{W}^l S_{t-1}^{l-1}}{\partial \boldsymbol{W}^l} + \frac{\partial R_{t-2}^l}{\partial \boldsymbol{W}^l}$$

$$\frac{\partial R_{t-1}^l}{\partial \boldsymbol{V}^l} \sim 0 + \frac{\partial \boldsymbol{V}^l R_{t-2}^l}{\partial \boldsymbol{V}^l}$$

Experimental Results

Pixel-by-pixel MNIST

Table 1: Accuracy of comparative models for the task of classifying pixel-pixel MNIST.

Types	Models	Cascade	Paras	Accuracy (%)
CNNs	CNN-SVM (Niu and Suen, 2012)	–	–	98.79
	LeNet-5 (LeCun et al., 1998)	–	6.2×10^4	99.05
MP-based	FCN with MP neurons	size(*,40,10)	3.2×10^4	86.47
	FCN with MP neurons	size(*,150,10)	1.2×10^5	93.82
SNNs	SLAYER	size(*,300,10)	2.4×10^5	94.13
	BSNN	size(*,300,10)	3.2×10^5	96.65
Recurrent Networks	uRNN (Arjovsky et al., 2016)	size(*,150,10)	3.3×10^4	97.28
	CNN-RNN	size(*,150,10)	2.9×10^4	95.21
	CNN-LSTM	size(*,150,10)	8.2×10^4	98.66
Our Work	FT0	size(*,0,10)	4.0×10^2	92.87
	FT1	size(*,150,10)	3.0×10^4	**99.12**

Univariate Time Series Forecasting- Yancheng Automobile Registration

Table 2: MSE and settings of comparative models for the task of forecasting Yancheng Automobile Registration records.

Types	Models	Settings	Paras	MSE (10^5)
Statistical Models	ARIMA	$(p, d, q) = (6, 1, 3)$	–	84.5129
	MAR (Won and Gray, 2013)	–	–	92.6458
	AGP (Blaauw et al., 2011)	–	–	41.0147
	KNNs (Yang et al., 2011)	$(K, w) = (1, 1)$	–	31.2573
Neural Networks	FCN with MP neurons	size(5,50,1)	200	19.5360
	FCN with MP neurons	size(5,500,1)	3.0×10^3	19.5360
	NARXnet (Guzman et al., 2017)	size(5,50,1)	–	20.2631
	RNN (Gers and Schmidhuber, 2000)	size(5,50,1)	2.8×10^2	18.0729
	LSTM (Hochreiter and Schmidhuber, 1997)	size(5,50,1)	8.3×10^2	10.7250
	LSTNet (Lai et al., 2018)	size(5,64,1)	9.7×10^3	8.4176
Our Work	FT0	size(5,0,1)	< 100	16.4721
	FT1	size(5,50,1)	2.9×10^3	**4.5067**

Multivariate Time Series Forecasting - Traffic Prediction on HDUK

Table 3: MSE and confusion accuracy of comparative models for the task of forecasting HDUK.

Data sets	Models & Settings / Evaluation(%)	NARXnet □	RNN △	LSTM △	LSTNet △	FT0 ◇	FT1 △
A1	MSE	0.0469	0.1499	0.0262	0.0247	0.1169	**0.0221**
	TPR	97.20	97.20	98.13	98.13	96.20	**99.07**
	TNR	95.29	91.74	96.47	97.41	94.12	**97.65**
A1033	MSE	0.1584	0.1716	0.1397	0.1401	0.1372	**0.1119**
	TPR	88.51	93.10	94.25	94.11	94.11	**96.55**
	TNR	91.43	93.33	92.38	92.25	92.38	**97.14**
A11	MSE	0.1754	0.1770	0.1725	0.1690	0.1755	**0.1651**
	TPR	97.06	96.08	97.06	97.06	97.06	**99.02**
	TNR	95.56	91.11	93.33	95.93	**96.67**	94.44

□ denotes a size(*,0,1) cascade structure and iterates 100 times;
△ denotes size(*,100,1) cascade structure and iterates 100 times;
◇ indicates a network configuration with 100 recurrent neurons and 32-dimensional convolution layer and iterates 100 epochs.