脉冲神经网络的训练

2023 8.29 PM
Day-07

Training spiking neural networks

王超名
chao.brain@qq.com

目录 CONTENTS

北京大学 PEKING UNIVERSITY

Q&A.
神经元的 bursting ~~的好处~~ 计算功能? 计算精度?
几百Hz 抑制它

Q GIF和LIF 区别?
为什么GIF可以训练好 LIF就不行
· 有些burst存在脑皮层同步化之类，在觉醒和注意力之类的发挥作用.

01 Introduction

类脑计算

So why should I care about SNNs?

- **Neuroscience**
- SNNs are what the brain does. If we want to fully understand the brain we need to understand SNNs.
- **Intelligence**
- They're the basis of the only known system for general intelligence.
- **Intellectual challenge**
- Find a way to think about hybrid continuous / discrete system, and computation based on this
- **The coming revolution**
- Exciting times for SNNs with new ways to train them (will talk about it in second half)
- **Neuromorphic hardware**
- Low power consumption.
- **Advantages of computation with SNNs?**
- Fast computation / rapid decision-making
- Multiplexing

大脑. spike train
外界 变化

Basic Elements of Training SNNs

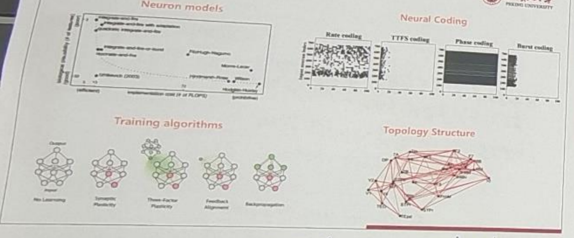
北京大学 PEKING UNIVERSITY

Neuron models

Neural Coding

Training algorithms

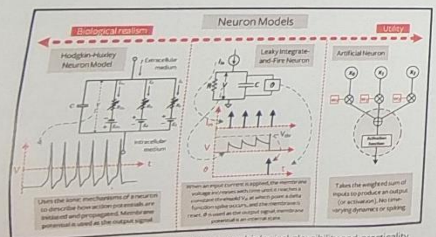
Topology Structure



02 Trainable Neuron Models

类脑计算: LIF model及其延伸

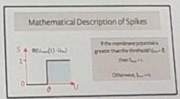
The spectrum of neuron models



The LIF neuron sits in the sweet spot between biological plausibility and practicality.

Non-differentiable problem

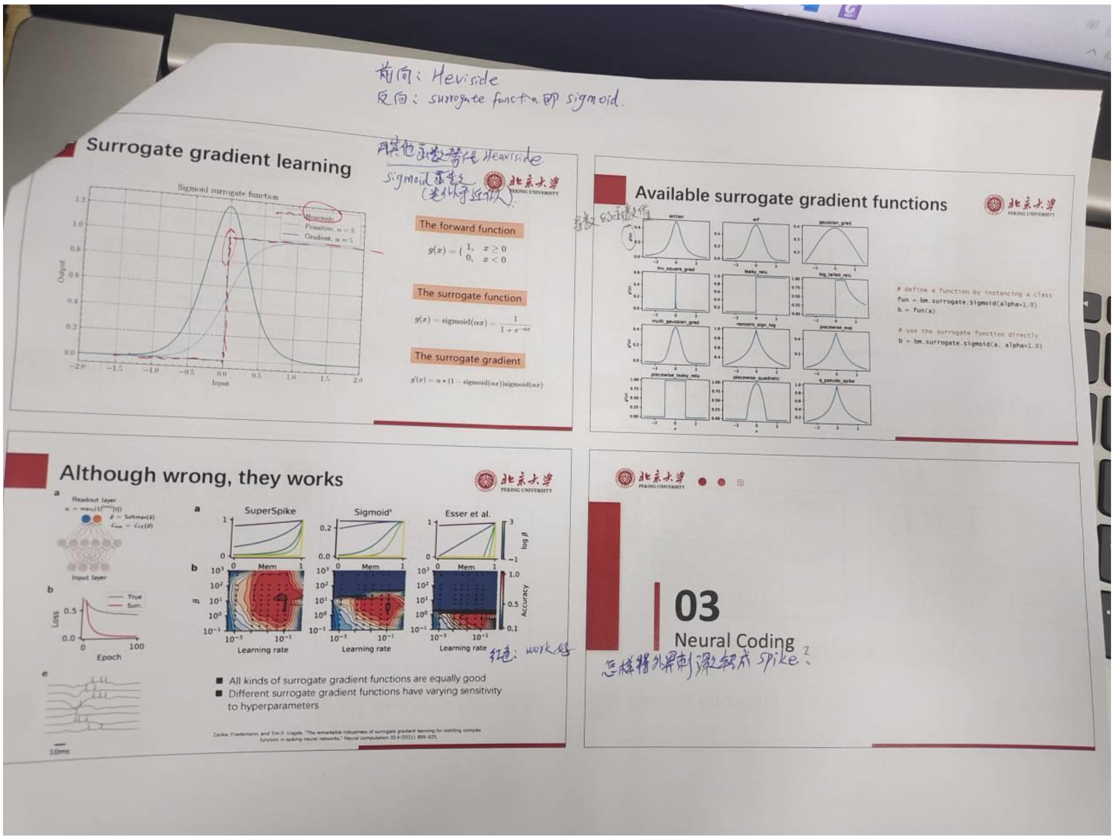
- The discharge process is an **non-differentiable** process
$$s(t) = \theta(U_{mem}(t) - U_{thr})$$
- $\theta(x) = \begin{cases} 1, & x \geq 0 \\ 0, & x < 0 \end{cases}$ is a **Heaviside step function**
- The derivative of the Heaviside function $\theta(x)$ is the **Delta function**
$$\delta(x) = \begin{cases} +\infty, & x = 0 \\ 0, & x \neq 0 \end{cases}$$
- Unable to use $\delta(x)$ to calculate during **backpropagation**

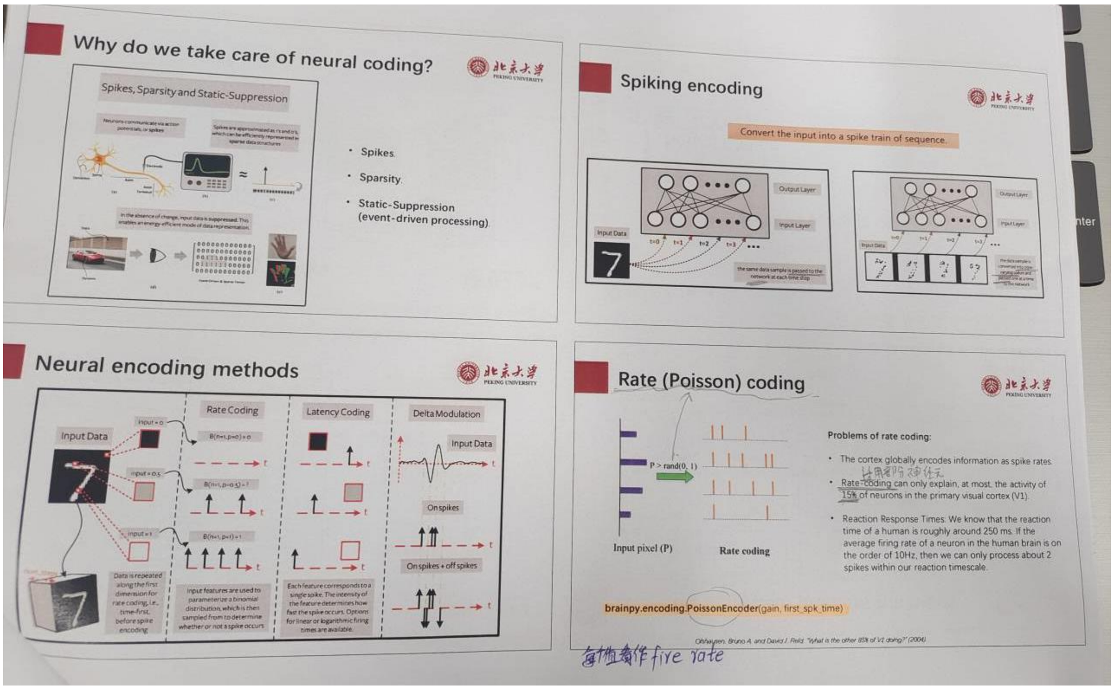


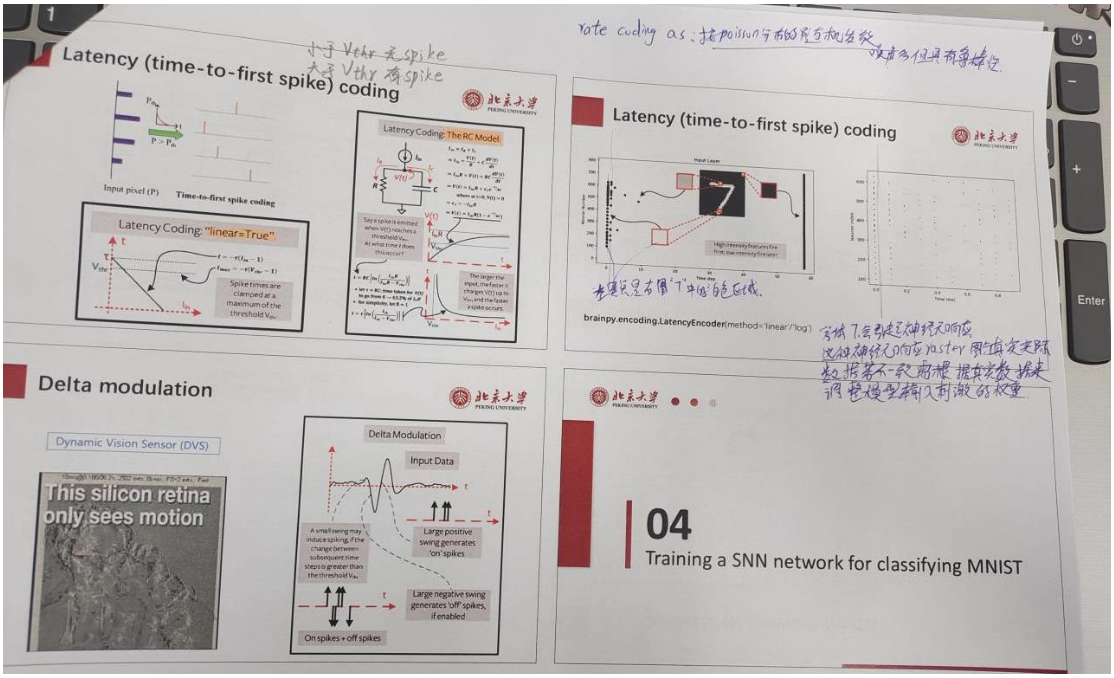
$$\frac{\partial C}{\partial W} = \frac{\partial C}{\partial S} \underbrace{\frac{\partial S}{\partial U}}_{\{0,\infty\}} \frac{\partial U}{\partial I} \frac{\partial I}{\partial W}$$ 链式法则

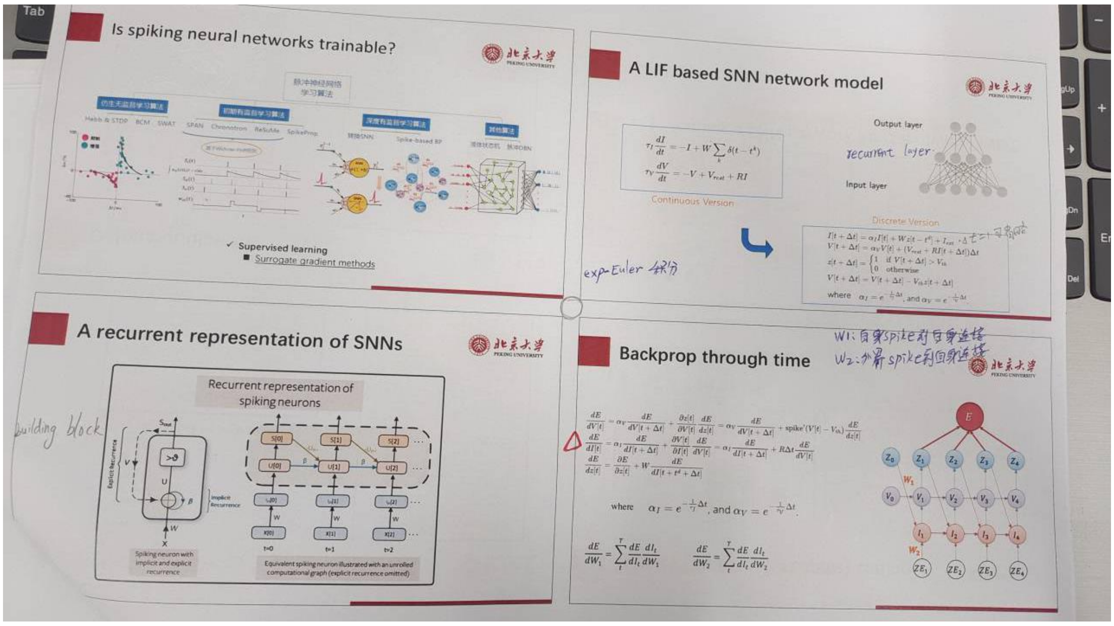
spike对膜电位

This is the problem!


前向：Heviside
反向：surrogate function 即 sigmoid.
Surrogate gradient learning
用其他函数替代 Heaviside
sigmoid 函数是 (类似于近似)
Sigmoid surrogate function
Heaviside
Primitive, α = 5
Gradient, α = 5
Output
Input
The forward function
$g(x) = \begin{cases} 1, & x \ge 0 \\ 0, & x < 0 \end{cases}$
The surrogate function
$g(x) = \mathrm{sigmoid}(\alpha x) = \frac{1}{1+e^{-\alpha x}}$
The surrogate gradient
$g'(x) = \alpha * (1 - \mathrm{sigmoid}(\alpha x))\mathrm{sigmoid}(\alpha x)$
Available surrogate gradient functions
# define a function by instancing a class
fun = bm.surrogate.Sigmoid(alpha=1.0)
b = fun(a)
# use the surrogate function directly
b = bm.surrogate.sigmoid(a, alpha=1.0)
Although wrong, they works
Readout layer
Input layer
True
Surr.
Loss
Epoch
SuperSpike
Sigmoid'
Esser et al.
Mem
Learning rate
Accuracy
红色：work好
All kinds of surrogate gradient functions are equally good
Different surrogate gradient functions have varying sensitivity to hyperparameters
03
Neural Coding
怎样将外界刺激变成 spike？


Why do we take care of neural coding?
Spikes, Sparsity and Static-Suppression
Spikes
Sparsity
Static-Suppression (event-driven processing)
Spiking encoding
Convert the input into a spike train of sequence.
Output Layer
Input Layer
Input Data
the same data sample is passed to the network at each time step
Neural encoding methods
Input Data
Rate Coding
Latency Coding
Delta Modulation
On spikes
On spikes + off spikes
Rate (Poisson) coding
P > rand(0, 1)
Input pixel (P)
Rate coding
Problems of rate coding:
The cortex globally encodes information as spike rates.
Rate-coding can only explain, at most, the activity of 15% of neurons in the primary visual cortex (V1).
Reaction Response Times: We know that the reaction time of a human is roughly around 250 ms. If the average firing rate of a neuron in the human brain is on the order of 10Hz, then we can only process about 2 spikes within our reaction timescale.
brainpy.encoding.PoissonEncoder(gain, first_spk_time)
Olshausen, Bruno A. and David J. Field. "What is the other 85% of V1 doing?" (2004)
每个值看作 fire rate


rate coding as: 按poisson分布的随机发放
小于 Vthr 无spike
大于 Vthr 有spike
Latency (time-to-first spike) coding
北京大学
PEKING UNIVERSITY
Input pixel (P)
Time-to-first spike coding
Latency Coding: "linear=True"
Spike times are clamped at a maximum of the threshold
Latency Coding: The RC Model
Latency (time-to-first spike) coding
Input Layer
brainpy.encoding.LatencyEncoder(method='linear'/'log')
Delta modulation
Dynamic Vision Sensor (DVS)
This silicon retina only sees motion
Delta Modulation
Input Data
Large positive swing generates 'on' spikes
Large negative swing generates 'off' spikes, if enabled
On spikes + off spikes
04
Training a SNN network for classifying MNIST


Is spiking neural networks trainable?
北京大学
脉冲神经网络学习算法
Supervised learning
Surrogate gradient methods
A LIF based SNN network model
Output layer
recurrent layer
Input layer
Continuous Version
Discrete Version
exp-Euler 积分
A recurrent representation of SNNs
Recurrent representation of spiking neurons
building block
Spiking neuron with implicit and explicit recurrence
Equivalent spiking neuron illustrated with an unrolled computational graph (explicit recurrence omitted)
Backprop through time
where