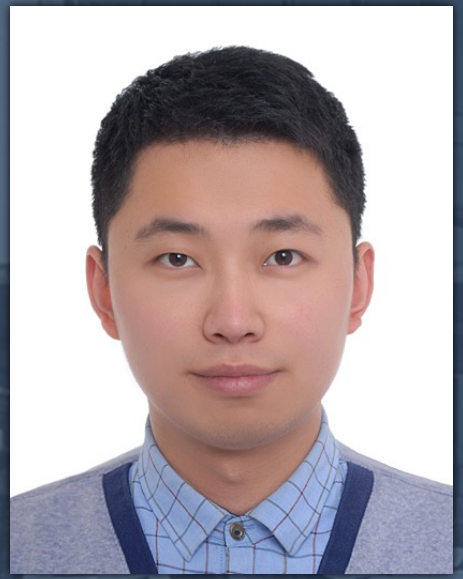
受智能启发的计算理论、方法与应用



助理教授，博士生导师
LAMDA 成员
Email: zhangsq@lamda.nju.edu.cn
zhangsq@nju.edu.cn (招生联络邮箱)
Web: https://www.lamda.nju.edu.cn/zhangsq/

智能科学与技术学院
School of Intelligence Science and Technology

Date: 2025.07.03

张绍群 助理教授、博导、LAMDA 成员

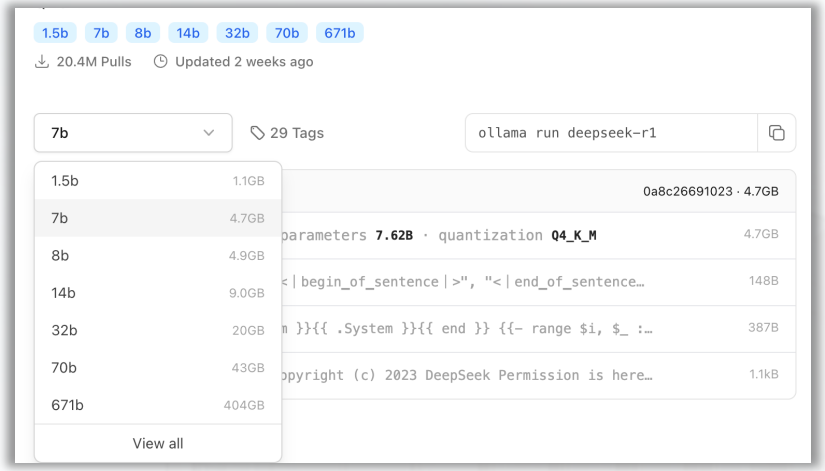
南京大学 NANJING UNIVERSITY 智能科学与技术学院 School of Intelligence Science and Technology

研究方向：机器学习与数据挖掘

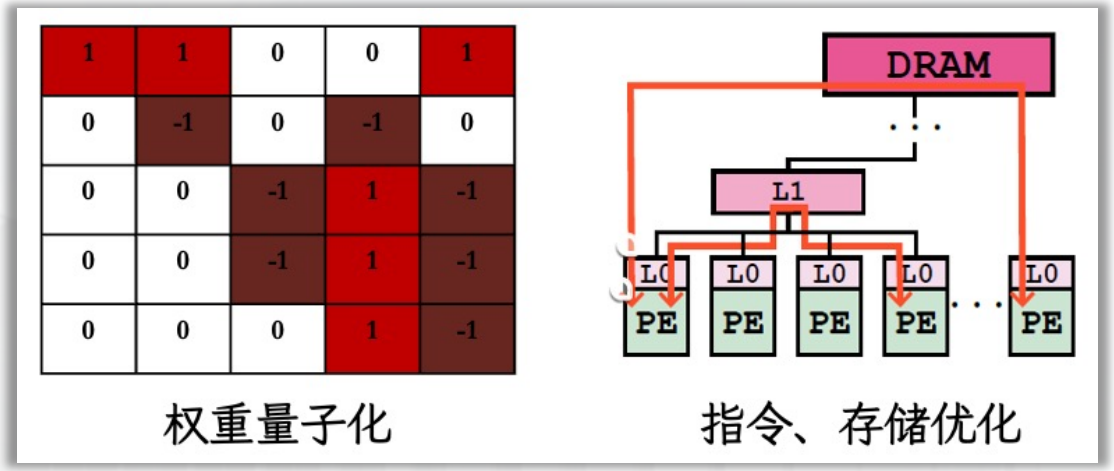
- **大模型的轻量化算法及理论 —— 高效低耗**
  - 在线检索推荐
  - 边缘计算、类脑计算等
- **时间序列分析 —— 时空大模型**
  - 工业自动化场景
  - 神经学、神经医学信号处理等
- **理论：机器学习理论，尤其是深度学习和大模型理论**
  - 表示和泛化理论 (Expressivity and Generalization)
  - 不确定性分析 (Uncertainty Analysis)
  - 知识与数据双驱动的反问题求解 (Inverse Problem)

个人主页

1. 大模型的轻量化算法及理论

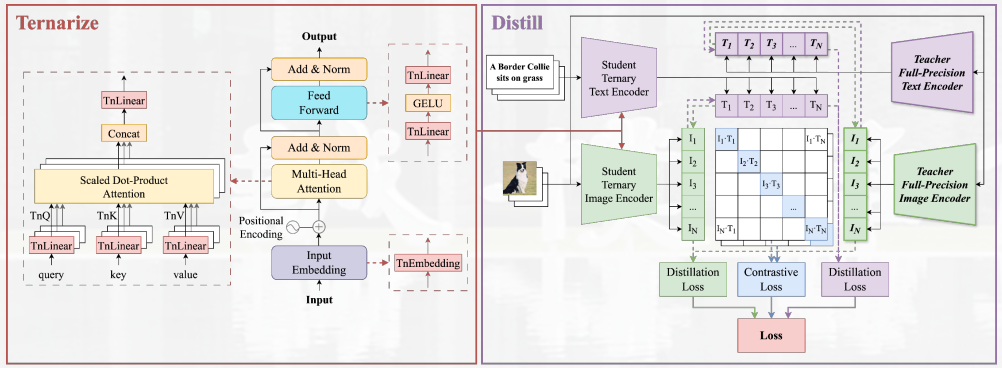


大模型“阉割”参数量 NO!

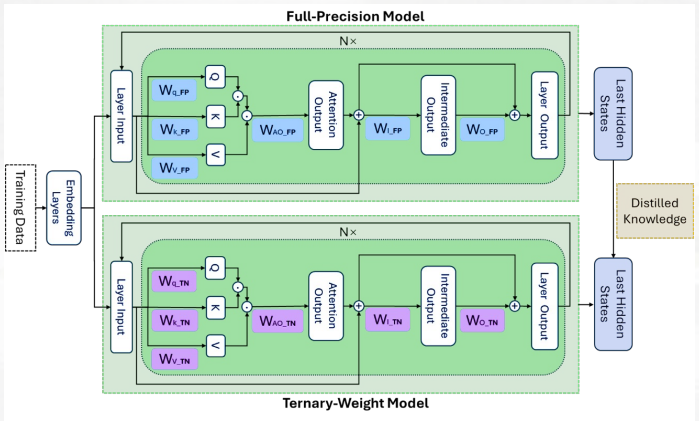


大模型压缩：高效低耗计算 Yes!

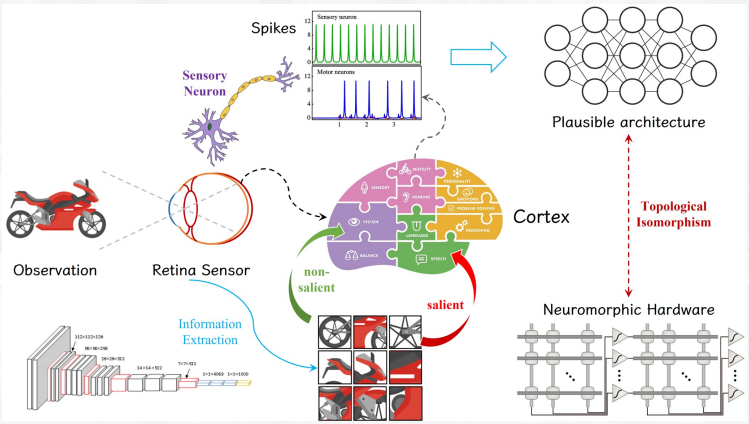
Zero-shot Multimodal Classification and Retrieval



Low-bit Embedding



Edge Computing



2. 时空大模型

Data ≈ 100 B

Comprehensiveness

Scaling Law > 2 B

工业

金融

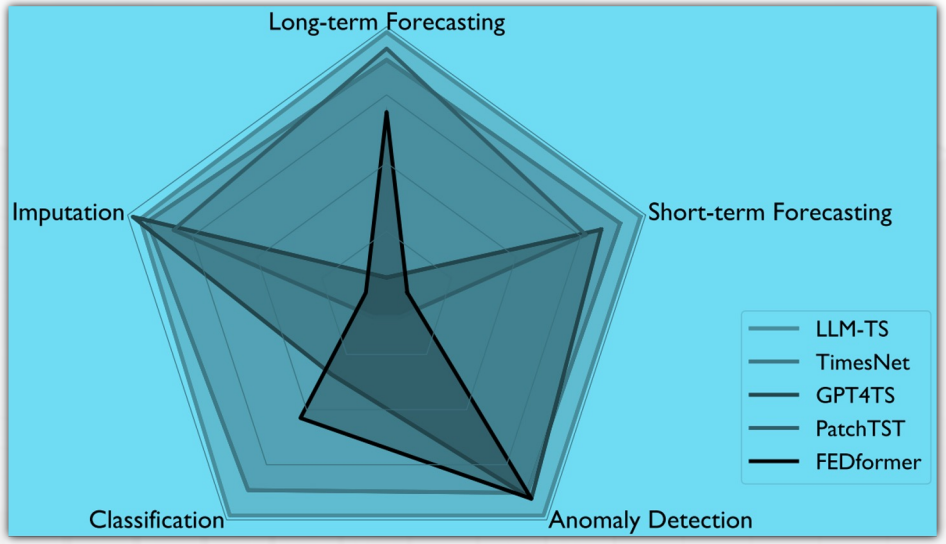
交通

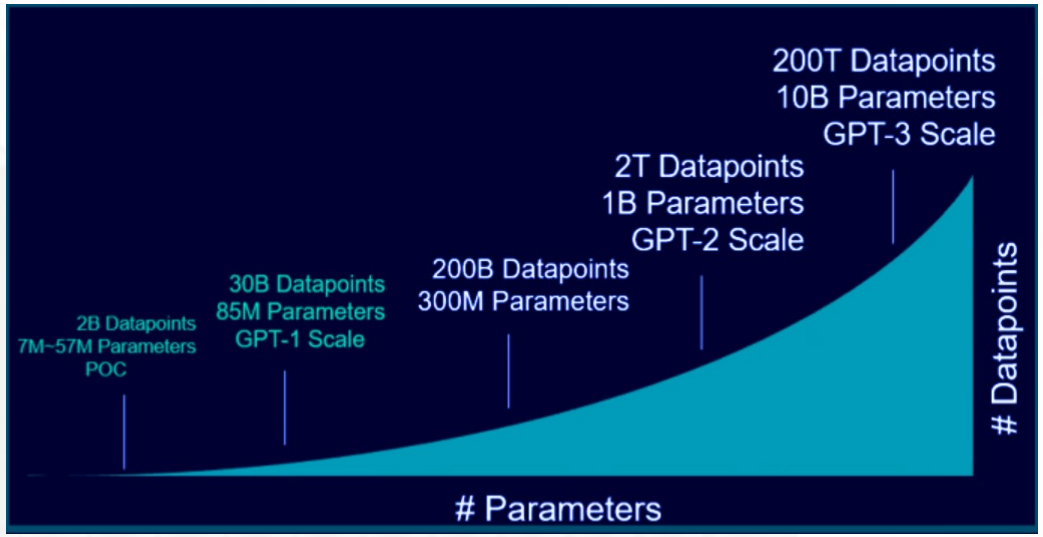
能源

天气

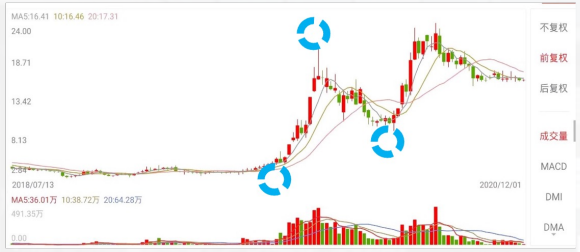
医疗

其他

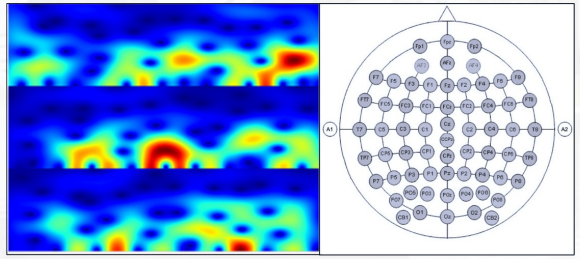




应用领域

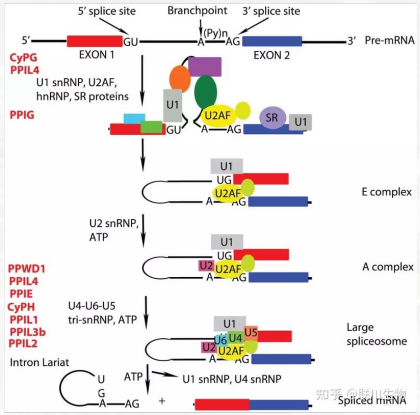


量化交易



神经信号处理

基因剪切

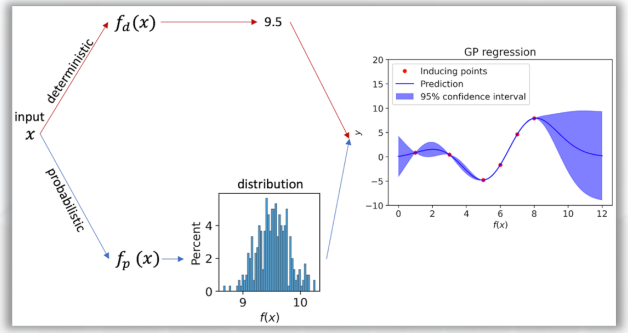


能源调控

3. 机器学习理论

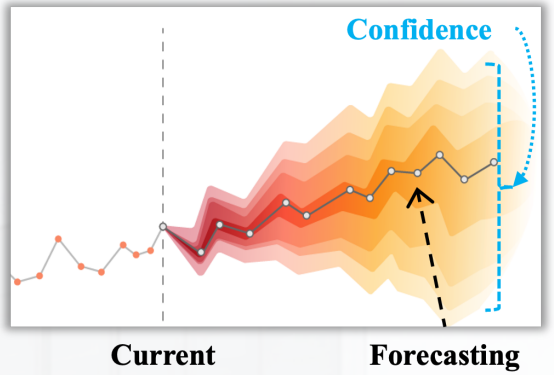
Uncertainty Analysis
不确定分析

(towards 幻觉和一致性等问题)

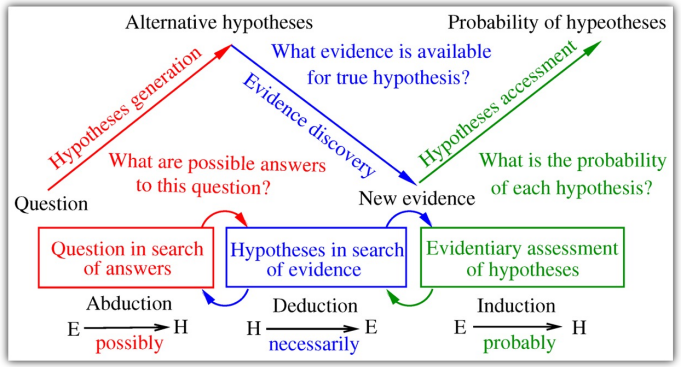


Quantification & Optimization

Make Decision with Confidence



Predictable PAC Learning Theory
可预测学习理论



Generalization bound with sequential Rademacher complexity

$$\mathbb{L}(\hbar) \leq \widehat{\mathbb{L}}_p(\hbar) + \boxed{\mathfrak{R}_p} + C_\ell \|\boldsymbol{\gamma}\|_2 \sqrt{2 \log \left(\frac{\boxed{\mathbb{E}_{\boldsymbol{\sigma}} \left[\mathcal{M}_{\boldsymbol{e}}(\chi, \mathcal{H}, \boldsymbol{\sigma}) \right]}}{\delta} \right)}$$

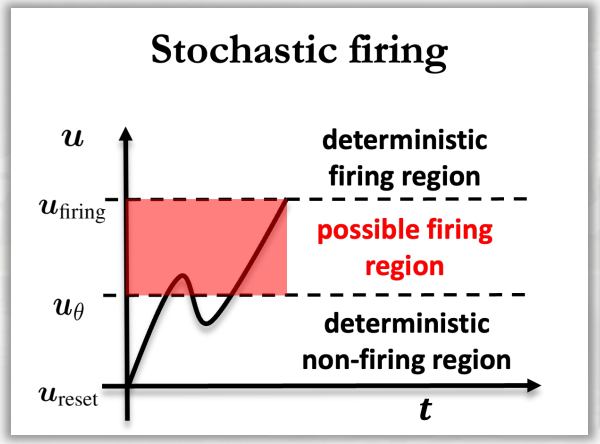
Discrepancy, estimated from Data

Sequential complexity, related to non-stationarity

where

$$\mathfrak{R}_p = \sup_{\hbar \in \mathcal{H}} \left[\mathbb{E}_{e_{T+1} \in \hat{\mathbf{E}}} \left[\ell(\hbar(\boldsymbol{e}_{-(T+1)}), e_{T+1}) \right] - \frac{1}{T} \left(\sum_{t=2}^{T} \gamma_t \mathbb{E}_{e_t \in \mathbf{E}} \left[\ell\left(\hbar(\boldsymbol{e}_{-t}), e_t\right) \right] + C \right) \right]$$

Bio-inspired Computing Theory
类脑计算理论



证明了首个关于脉冲神经网络的万有逼近定理、结构稳定性、泛化性

Theorem 12 *Let $\mathcal{F}_{\mathcal{W}}^{one}$ denote the function space of L-layer stochastic neurons. If $u_{reset} = 0$, $\|\mathbf{W}^l\|_2 \leq C_l$ for $l \in [L]$, and $\|\mathbf{X}\|_2 \leq C_X$ for $\mathbf{X} \in \mathcal{X}$, we have*

$$\mathfrak{R}_n(\mathscr{L} \circ \mathcal{F}_{\mathcal{W}}) \leq C_n \, \mathfrak{R}_n(\mathcal{F}_{\mathcal{W}}) \leq \frac{(C_n)^L C_X}{\sqrt{n}} \boxed{\left(\prod_{l \in [L]} C_l \right) (p_{max})^{(L+1)/2}}, \tag{42}$$

where $\boxed{p_{max} = \max_{i \in [n], l \in [L]} \{1 - \max\{p^{(l,i)}, p_\theta\}\} \in (0,1)}$ *and C_n is a universal constant.*

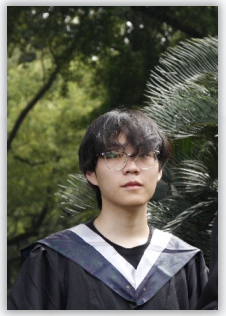
合作与文化

南京大学 NANJING UNIVERSITY

智能科学与技术学院 School of Intelligence Science and Technology

Teams & Cooperation

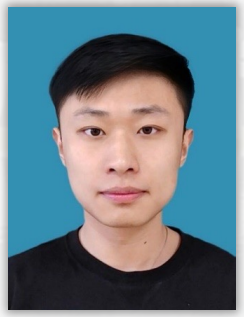











Join us!

● ● ●

Academic & Cooperation

Microsoft

SIEMENS

中国移动 China Mobile

JS NHC PRC

THE AFFILIATED DRUM TOWER HOSPITAL OF 1892 鼓楼医院 NANJING UNIVERSITY MEDICAL SCHOOL

Culture