



**NANYANG
TECHNOLOGICAL
UNIVERSITY**

Extreme Learning Machines (ELM)

超限学习机

- 筑梦智能物联网时代的普适学习和普适智能

Guang-Bin Huang (黄广斌)

Professor of School of Electrical and Electronic Engineering

Nanyang Technological University, Singapore

(新加坡南洋理工大学电子电气工程学院 教授)

2016中国软件大会 中国·北京 2016年12月22-23日

历史进程：人工智能和机器学习的三浪发展

1950s-1980s: 预热

特点：计算机能力有限，没有有效算法，没有足够多的数据

形势：华人创造性地把计算机翻译成**电脑**，是巧合、梦想还是预见？

1980s-2010: 研究驱动

特点：计算机功能足够强大，开发出许多算法，许多情况下没有足够多的数据

形势：少工业产业驱动，研究更多的来自研究人员和学术界的自发自我追求

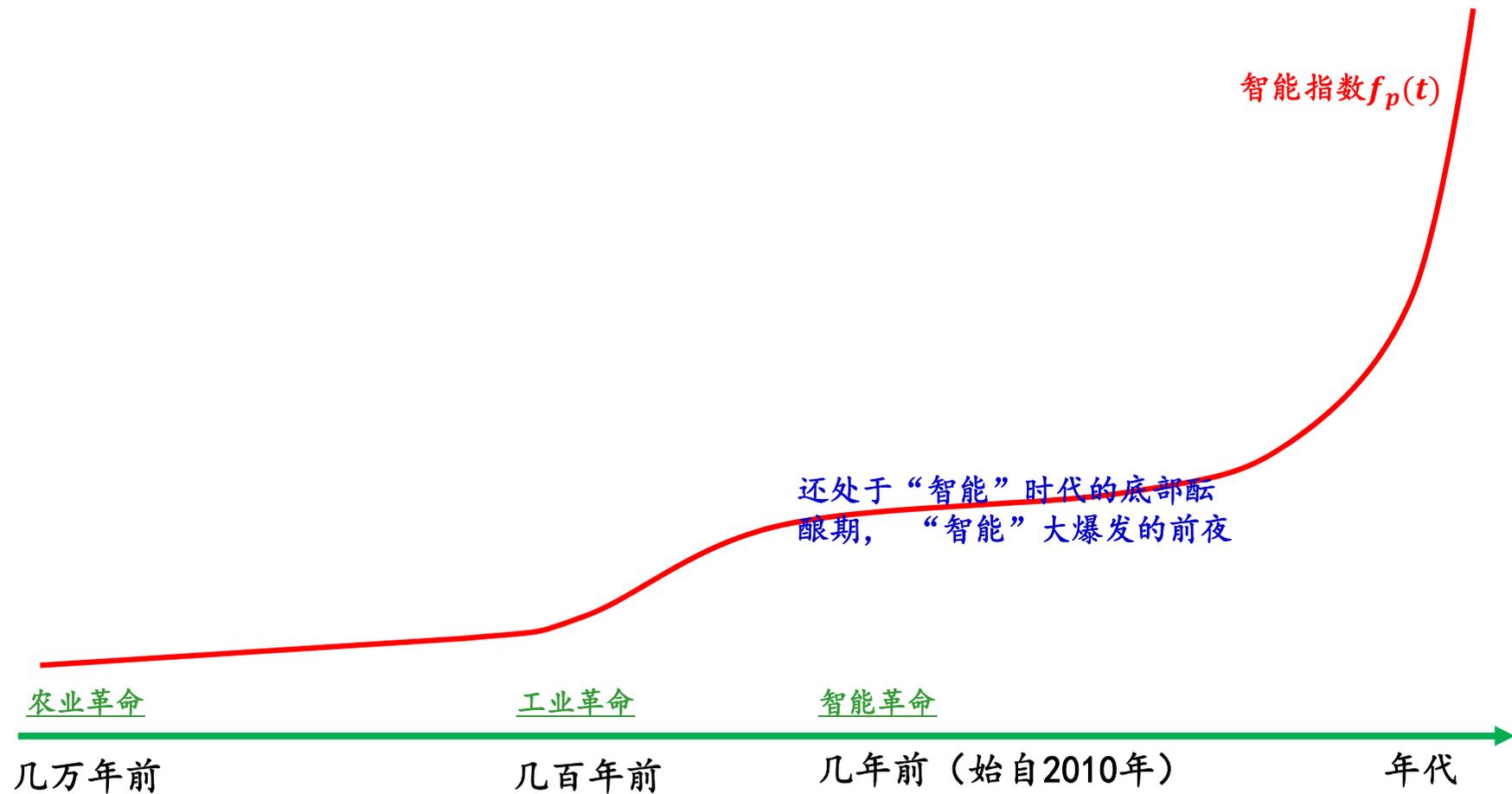
2010 - 现在: 数据/产业驱动

特点：计算机功能足够强大，处处是智能传感器和智能单元，数据超大，有效算法逐步推进

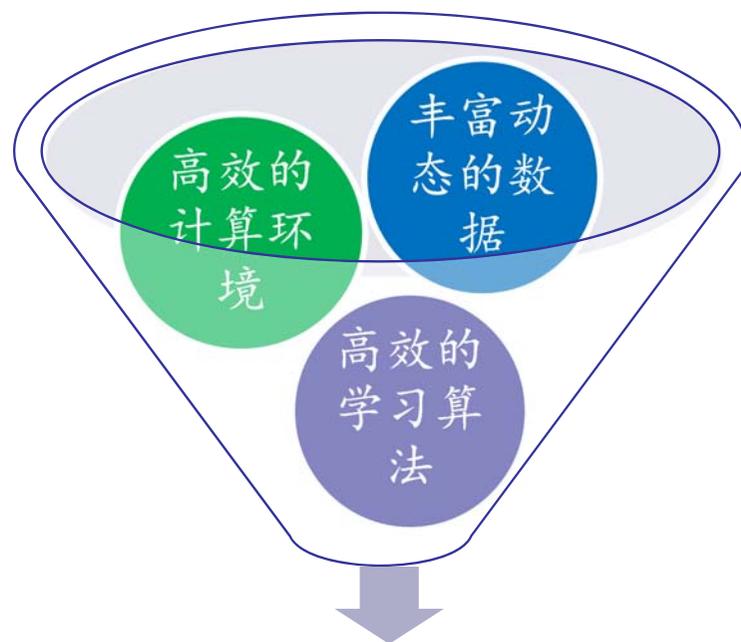
人类从此也变成了一个“超级智能传感器”

形势：不管人们信不信，人们只能愈来愈依赖人工智能和机器学习；相比于“农业革命”和“工业革命”，“智能革命”影响更大、更深远

历史进程：三大革命

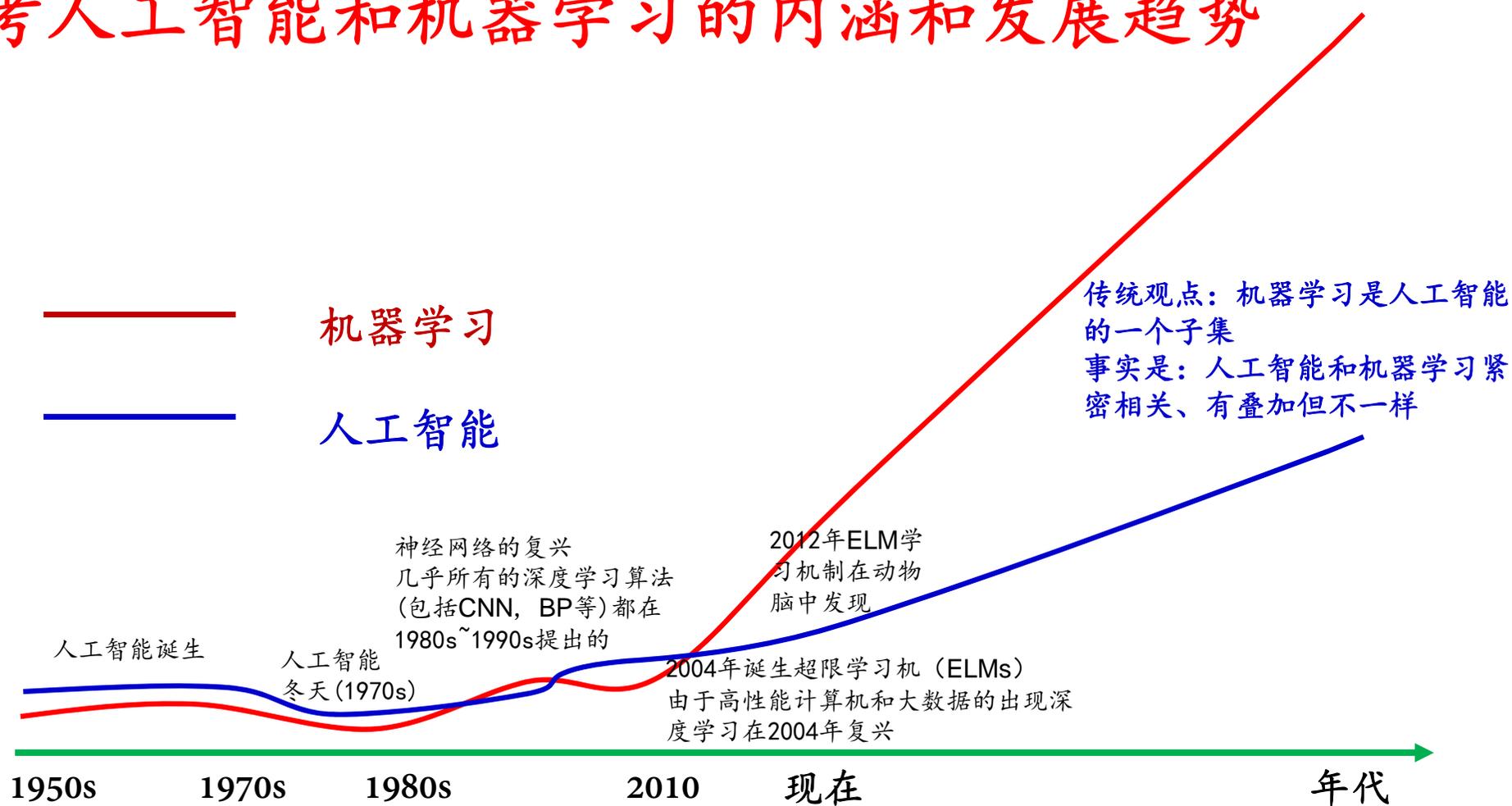


机器学习时代的必要条件

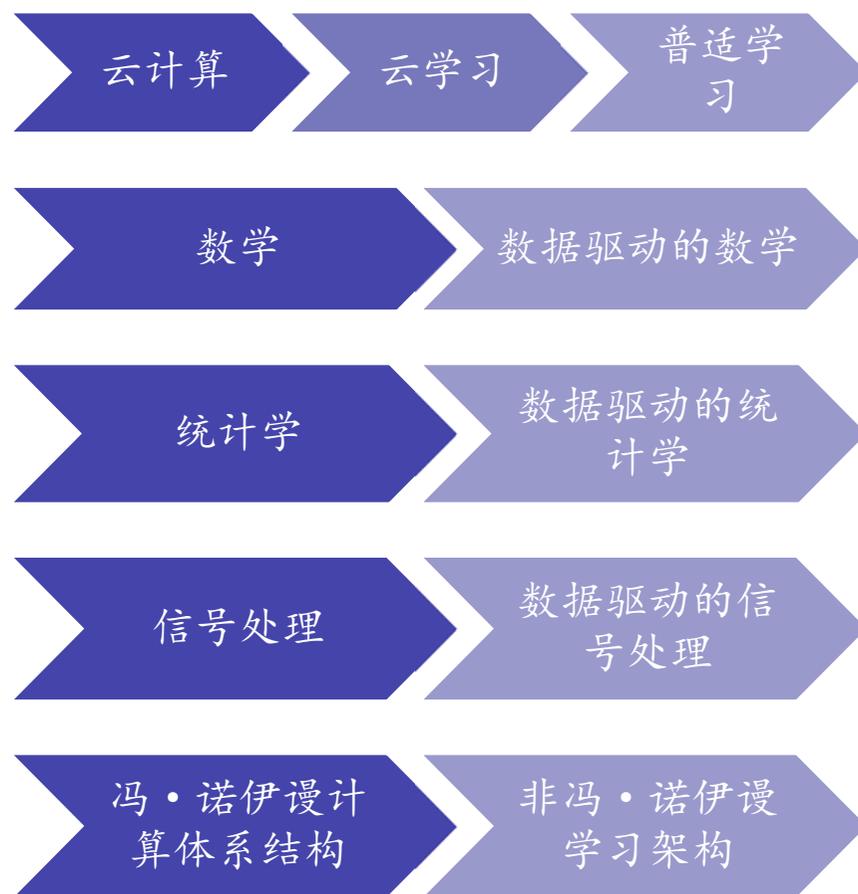


自从2010年开始这三个必要条件已经同时满足，虽然还处在及其初级阶段，还在不断发展提高，但这并不影响广泛应用和需求。

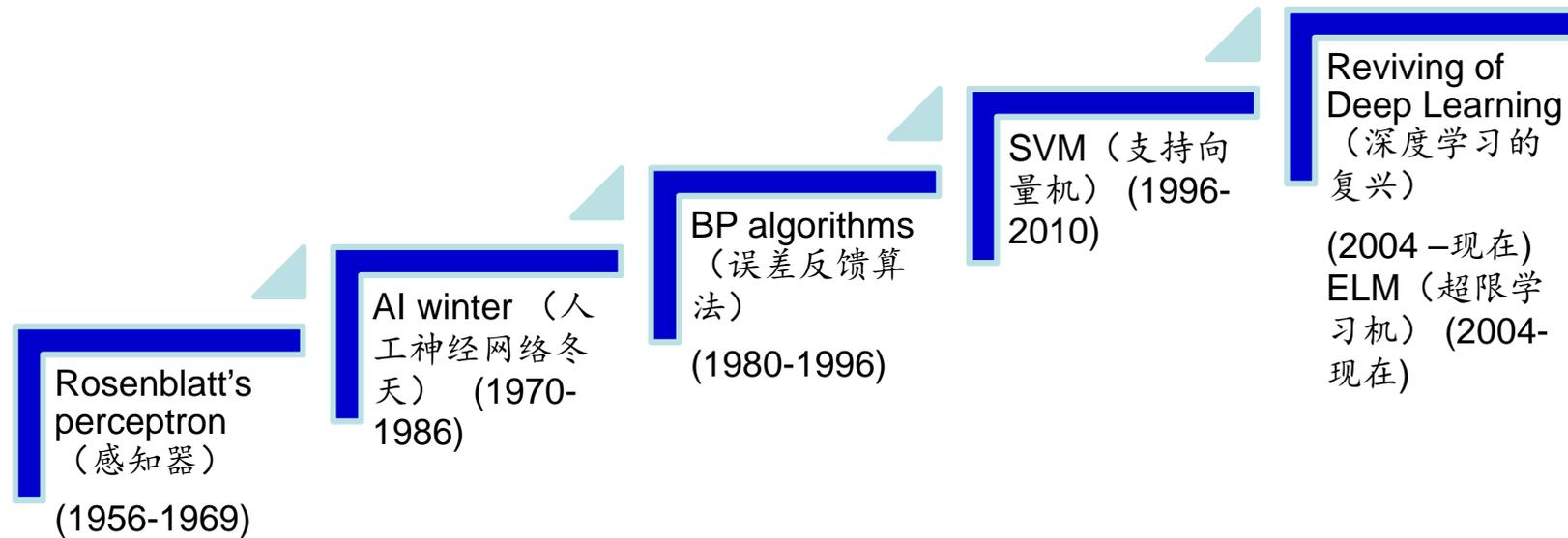
再思考人工智能和机器学习的内涵和发展趋势



历史进程：从机器学习工程到机器学习科学



人工神经网络历史的魔幻“15年”周期



It's time to move toward (Machine) Learning Science from Machine Learning Engineering

是时候从机器学习工程迈向机器学习科学

New turning point is coming, new machine learning techniques are on the way

新的转折点即将到来，新的机器学习和人工智能“火花”也许正在学界酝酿产生，学界和工业界的相互关系没有也不会改变：“新理论和新思想首先在学界酝酿，然后在产业界变‘现’（即：实施和形成生产力）”

人工神经网络的理论

• 人工神经网络应用的理论基础

- 万能逼近能力定理 (Universal Approximation capability) [Leshno 1993, Park and Sandberg 1991, Chen, et al 1995]: 任何连续目标函数可以用前馈神经网络以任意小的误差近似逼近。
- 分类能力定理 (Classification capability) [Huang, et al 2000]: 任何理论上可以分开的目标都可以用人工 (前馈) 神经网络加以分开。

• 人工神经网络学习的问题核心

- 这些理论都只是在理论上回答了**存在性**: 给定某种应用存在能提供所需解决能力的对应的人工神经网络结构。
- 但这两个定理没有能回答和提供学习的方法 (网络结构和算法) 。

M. Leshno, et al., "Multilayer feedforward networks with a nonpolynomial activation function can approximate any function," *Neural Networks*, vol. 6, pp. 861-867, 1993.

J. Park and I. W. Sandberg, "Universal approximation using radial-basis-function networks," *Neural Computation*, vol. 3, pp. 246-257, 1991.

G.-B. Huang, et al., "Classification ability of single hidden layer feedforward neural networks," *IEEE Trans. Neural Networks*, vol. 11, no. 3, pp. 799-801, May 2000.

人工神经网络和生物学习

深度学习 Deep Learning

对网络结构大小极其敏感

极其”痛苦”的人工”调参”

在并行和硬件实现上比较难

很难微型实时学习；需要大量的时间；很难实现多通道数据实时融合和决策系统实时同步

很难实现在线串行数据学习（sequential learning）和数据流学习（stream data learning）

大数据高精度；许多简单的应用也需要才能大数据实现

需要海量计算资源（通常几千几万个快速计算单元的并行运算）

学习速度极其低速，在学习精度上极其“贪婪”

总是先有应用，再去找寻相匹配的“个性化”强的学习算法

生物学习 Biological Learning

对生物神经网络模块大小不是苛求，稳定性很高（每个模块大小不等：几十~几万个神经元）

并没有什么“人”在“脑”中不时“调参”

并行和硬件实现是生物学习机制的天然实现模式

微型实时学习；天然实现多通道数据实时融合和决策系统实时同步

天然在线串行数据学习和数据流学习

用小数据解决许多复杂应用

低速的神经元并行学习

讲究高学习速度，高的准确率，但不过分“贪婪”于学习精度

总是先有普适的“脑”，再有应用





NANYANG
TECHNOLOGICAL
UNIVERSITY

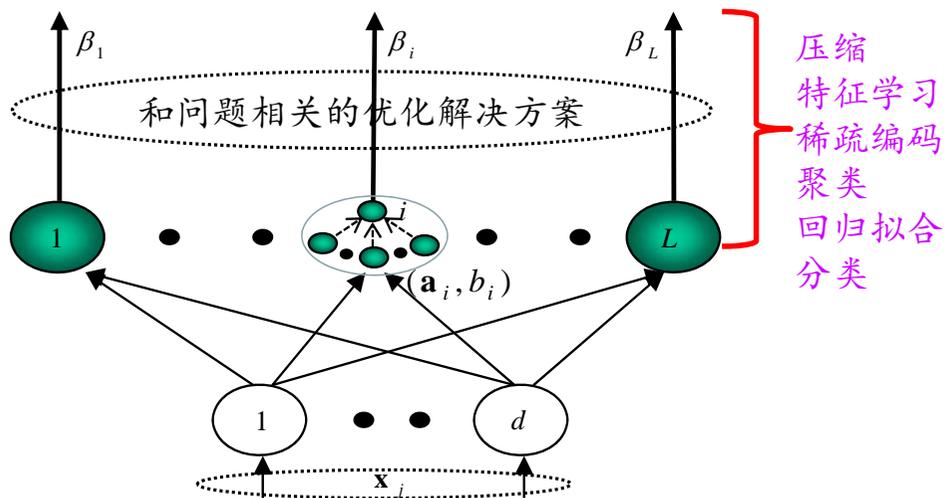
人工神经网络学习真的需要如此
“痛苦”和“耗时”？

让人工神经网络研发能变得轻松
愉快？

超限学习机 (Extreme Learning Machines, ELM) : 基本问题

- 面对如此多的神经网络结构，真的需要如此多的不同的对应神经网络算法？
 - 不同前馈网络结构（不仅限于Sigmoid/RBF网络）
 - sigmoid networks
 - RBF networks
 - polynomial networks
 - complex (domain) networks (复变域网络)
 - Fourier series (傅立叶变换)
 - wavelet networks (小波网络), etc
 - 多层网络（不仅限于多层网络，包括层次性网络）
- 生物脑（生物学习系统）中有几十到几百种不同神经元，（其数学模型甚至未知），问题是在生物学习中真的要调整各种生物神经元？

超限学习机 (ELM)



L 随机隐层节点 (几乎任何非线性阶段连续函数):
 $h_i(\mathbf{x}) = G_i(\mathbf{a}_i, b_i, \mathbf{x})$

虽然生物神经元的输出函数的数学模型未知, 但是几乎所有生物神经元可以看成是非线性阶段连续函数, 为ELM理论所覆盖。

ELM网络输出: $f_L(x) = \sum_{i=1}^L \beta_i G(\mathbf{a}_i, b_i, \mathbf{x})$

ELM特征映射: $h(x) = [G(\mathbf{a}_1, b_1, \mathbf{x}), \dots, G(\mathbf{a}_L, b_L, \mathbf{x})]$

隐层节点输出函数 (单节点):

Sigmoid: $G(\mathbf{a}_i, b_i, \mathbf{x}) = g(\mathbf{a}_i \cdot \mathbf{x} + b_i)$

RBF: $G(\mathbf{a}_i, b_i, \mathbf{x}) = g(b_i \|\mathbf{x} - \mathbf{a}_i\|)$

Fourier Series: $G(\mathbf{a}_i, b_i, \mathbf{x}) = \cos(\mathbf{a}_i \cdot \mathbf{x} + b_i)$

隐层复合节点输出函数 (子网络节点)

超限学习机 (ELM)

- **新的学习理论 (ELM学习理论)** - (机器或生物) 学习可以不需要调整隐层节点: 给定任何连续目标函数或可分类目标, 只要前馈神经的隐层节点是非线性阶段连续的, 神经网络无需调整隐层节点就能任意逼近目标连续函数或对分类目标加以分类, 首次理论证明SVM提供的是次优学习方法。 [Huang, et al 2004, 2006, 2007, 2012]
 - 新的学习理论不仅解决了神经网络存在性问题也给出了学习方法, 学习速度是可以达到深度学习和SVM的几万倍。
 - 我们首先提出和回答了生物学习的一个基本问题: “生物神经元在学习中真的需要调整吗?”
 - (机器或生物) 学习可以不需要调整隐层节点的三方面实现:
 - 上一代传给下一代
 - 从一个系统传给另外一个系统
 - 随机神经元/随机连接

G.-B. Huang, et al., "Universal approximation using incremental constructive feedforward networks with random hidden nodes," *IEEE Transactions on Neural Networks*, vol. 17, no. 4, pp. 879-892, 2006.

微博: weibo.com/elm201x; www.extreme-learning-machines.org

超限学习机 (ELM)

- **新的学习理论 (ELM学习理论)** - (机器或生物) 学习可以不需要调整隐层节点: 给定任何连续目标函数或可分类目标, 只要前馈神经的隐层节点是非线性阶段连续的, 神经网络无需调整隐层节点就能任意逼近目标连续函数或对分类目标加以分类, 首次理论证明SVM提供的是次优学习方法。 [Huang, et al 2004, 2006, 2007, 2012]
 - ELM理论在提出后约10年, 哈佛、哥伦比亚大学、斯坦福、MIT、IBM Watson、Georgia Tech等研究人员分别在2013年、2015年和2016年在老鼠的嗅觉系统、猴子的视觉系统和人的感知系统直接或间接的得到验证。
 - 作为ELM的一个特例 (傅立叶变换作为隐层输出), **Random Kitchen Sink (RKS)** 在Google、微软、斯坦福等广泛应用 [Rahimi and Recht 2009]。
 - 2016年IBM研究出基于ELM的第一个类生物神经元。
 - 随机神经元/随机连接也逐步在越来越多的深度学习中应用 (Google, Intel, Facebook, 微软等)。ELM提供了使用的理论基础。



O. Barak, et al, "The importance of mixed selectivity in complex cognitive tasks," *Nature*, vol.497, 2013

M. Rigotti, et al, *Journal of Neuroscience*, vol. 33, no. 9, 2013

R. I. Arriaga, et al. "Visual Categorization with Random Projection," *Neural Computation*, vol. 27, 2015

E. L Rich and J. D Wallis, "What stays the same in orbitofrontal cortex," *Nature Neuroscience*, vol. 19, no. 6, 2016

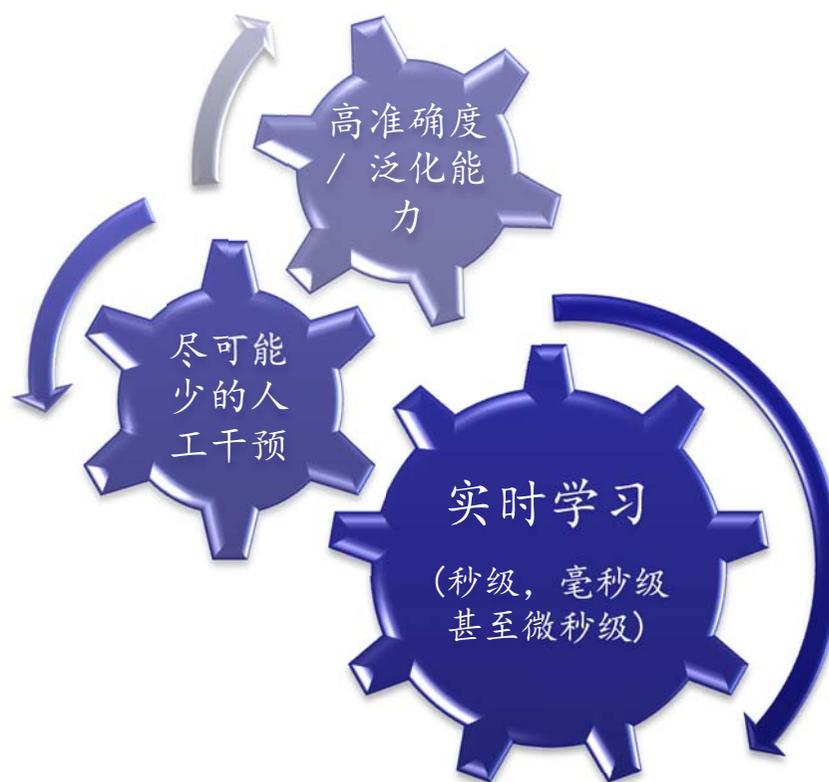
J. Xie and C. Padoa-Schioppa, "Neuronal remapping and circuit persistence in economic decisions," *Nature Neuroscience*, vol. 19, 2016

E. L Rich and J. D Wallis, "What stays the same in orbitofrontal cortex," *Nature Neuroscience*, vol. 19, no. 6, 2016

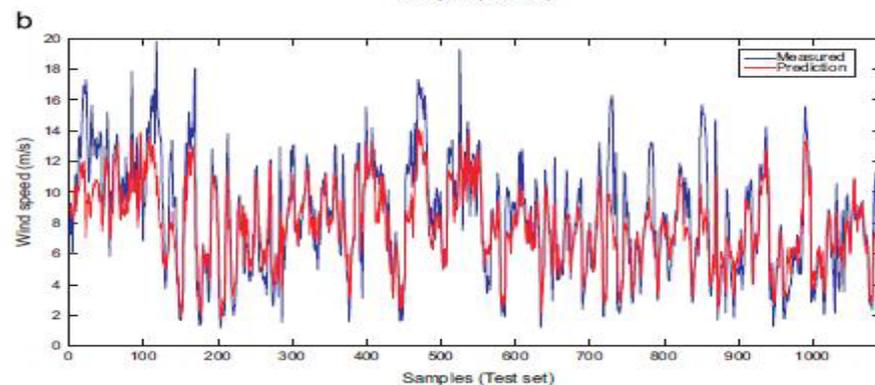
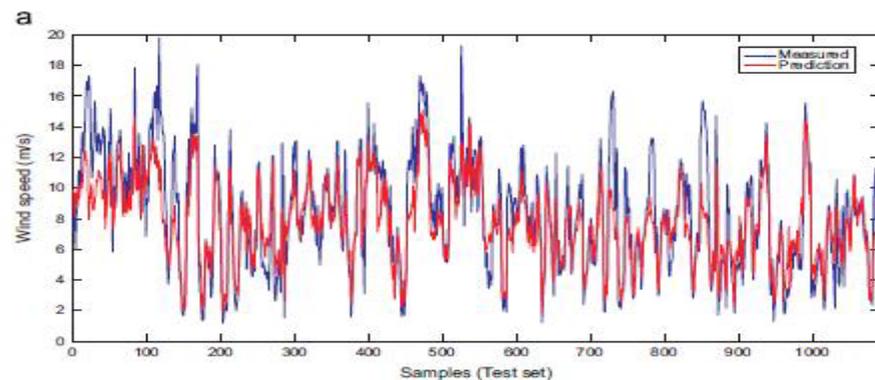
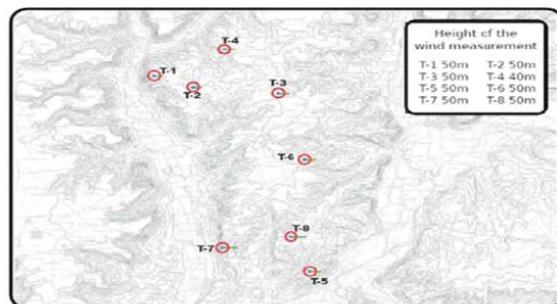
T. Tuma, et al, "Stochastic phase-change neurons," *Nature Nanotechnology*, vol. 11, 2016

微博: weibo.com/elm201x; www.extreme-learning-machines.org

ELM的核心考虑

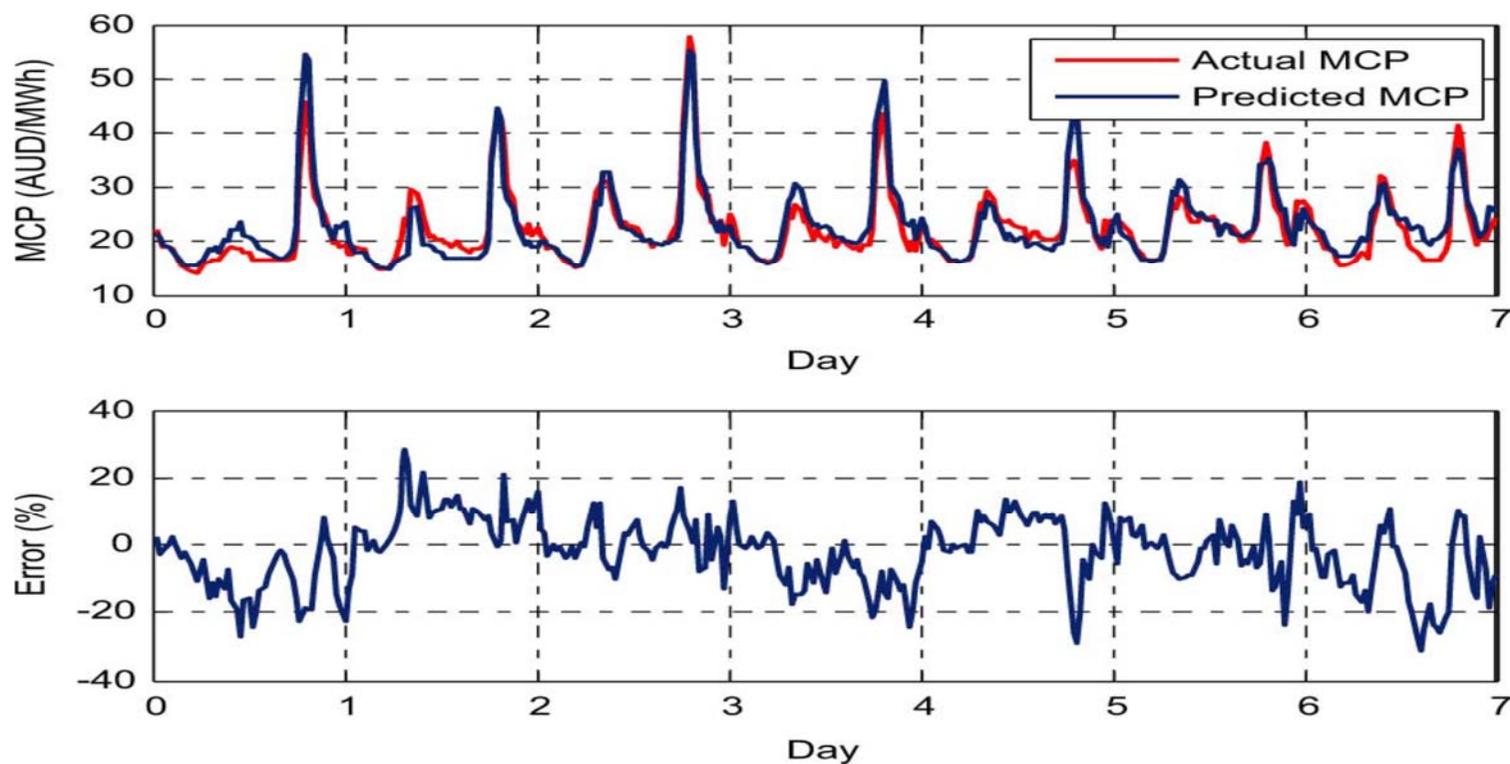


ELM应用：风力发电 - 西班牙



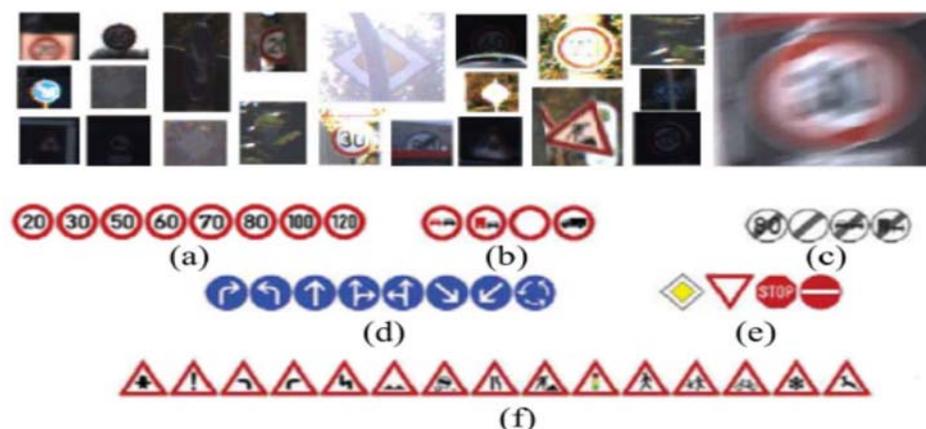
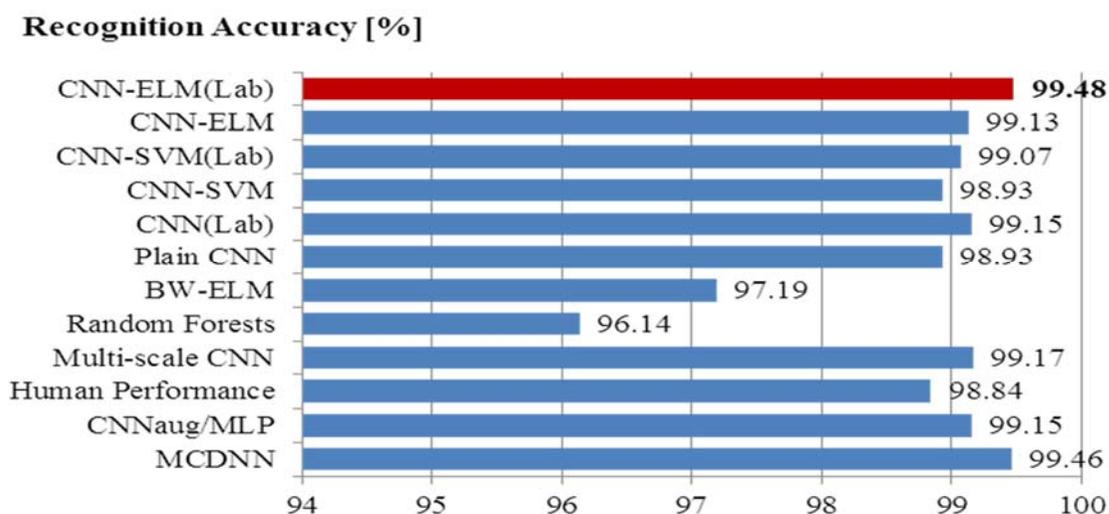
Situation of the wind measuring towers in Spain and within the eight wind farms. Wind speed prediction in tower 6 of the considered wind farm in Spain obtained by the ELM network (prediction using data from 7 towers). (a) Best prediction obtained and (b) worst prediction obtained. [Saavedra-Moreno, et al, 2013]

ELM应用：电价预测 - 澳洲



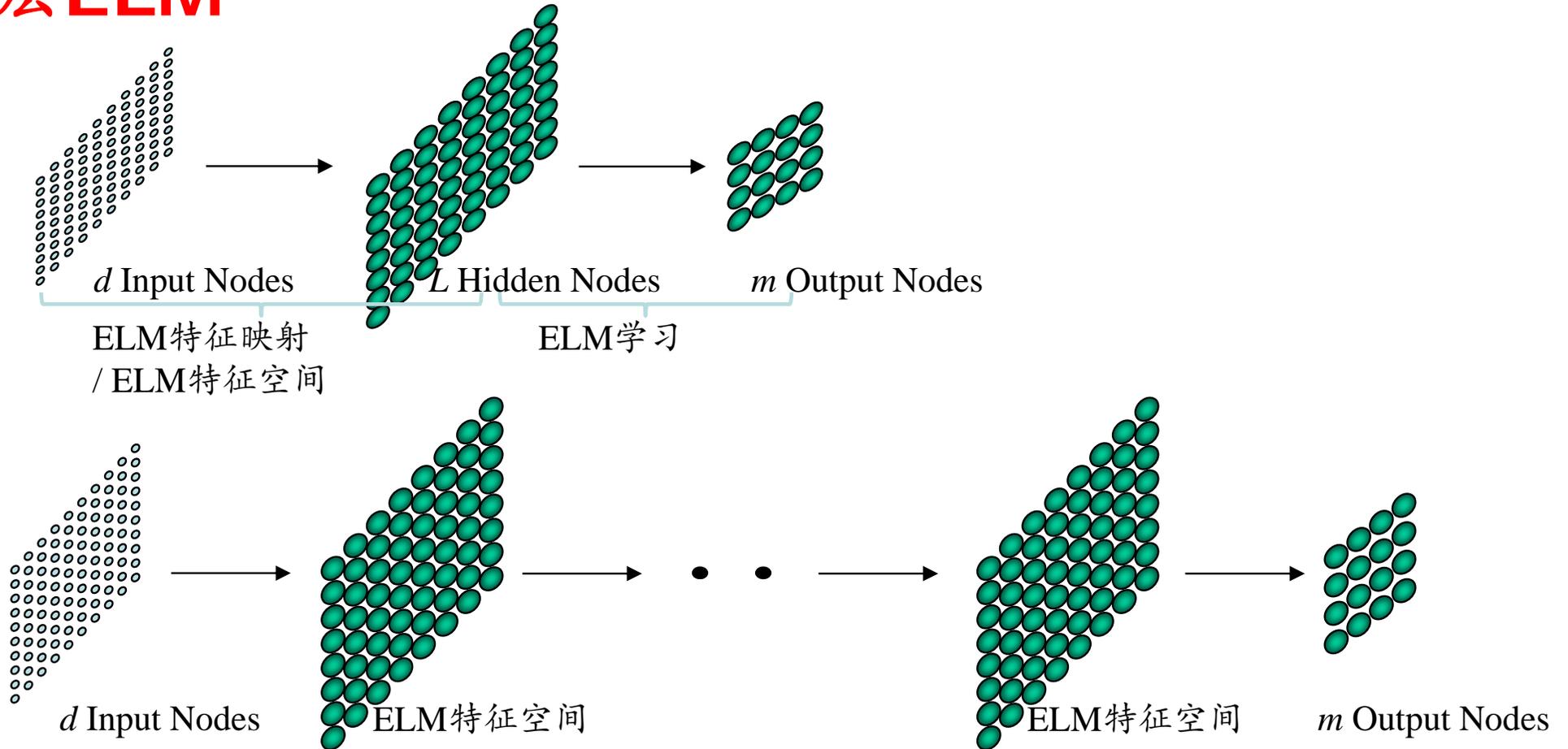
ELM应用：交通路牌识别

| 学习方式 (for GTSRB dataset: 39209 training data, 12630 testing data, 43 classes) | 测试精度 | 训练时间 |
|---|--------|--------------|
| HOG+ELM | 99.56% | 209秒 (普通PC) |
| CNN + ELM Based | 99.48% | 5小时 (普通PC) |
| MCDNN | 99.46% | 37小时 (GPU实现) |



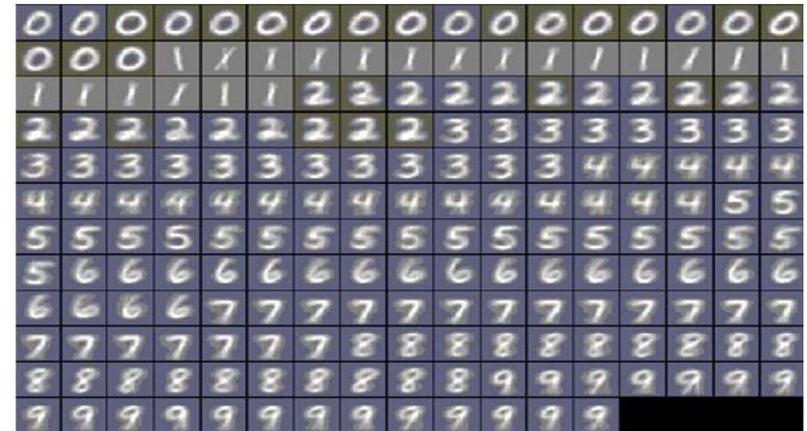
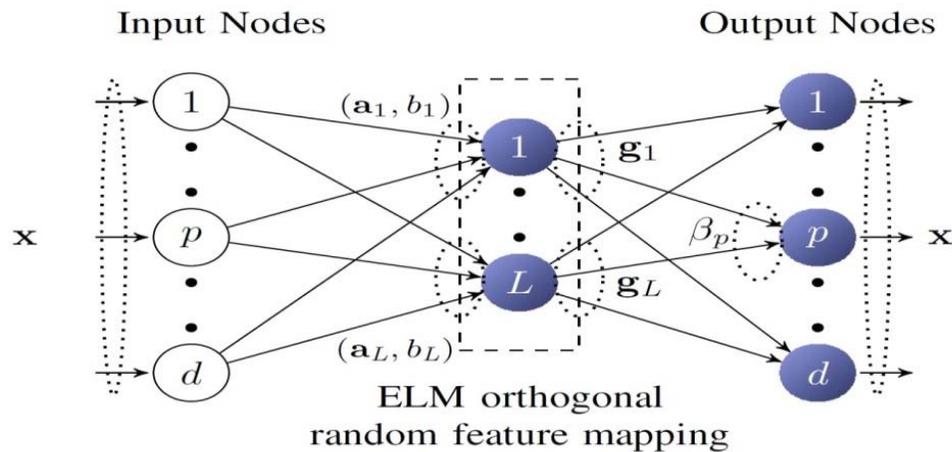
Z. Huang, Y. Yu, J. Gu, and H. Liu, "An Efficient Method for Traffic Sign Recognition Based on Extreme Learning Machine," *IEEE Transactions on Cybernetics*, 2016

多层ELM



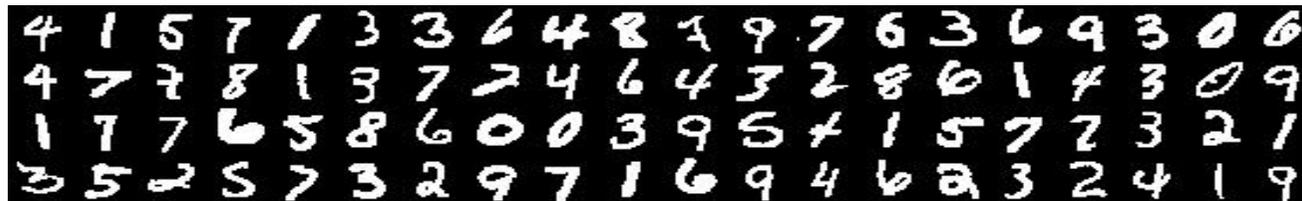
Different from Deep Learning, hidden neurons in multi-layer / hierarchical ELM (considered as a single whole ELM) are not required to be iteratively tuned

多层ELM应用：ELM自编码器(ELM-AE)



- $d > L$: Compressed Representation
- $d = L$: Equal Dimension Representation
- $d < L$: Sparse Representation

Features represented by the output weights of ELM-AE of MNIST OCR Datasets (with 60000 training samples and 10000 testing samples)



多层ELM应用：手写体识别

| 学习方式 | | 测试精度 | 训练时间 |
|------|--|------------|------------------------|
| ELM | H-ELM [Unpublished] | ~99.6% | 分钟(CPU) |
| | H-ELM [J. Tang, et al, 2015] | 99.14 | 281.37秒(CPU) |
| | Multi-Layer ELM (784-700-700-15000-10) [Huang, et al 2013] | 99.03±0.04 | 444.7秒(CPU) |
| 深度学习 | Deep Belief Networks (DBN) (748-500-500-2000-10) | 98.87 | 20580秒 (5.7小时, GPU) |
| | Deep Boltzmann Machines (DBM) (784-500-1000-10) | 99.05 | 68246秒 (19小时, GPU) |
| | Stacked Auto Encoders (SAE) | 98.6 | >17小时, GPU |
| | Stacked Denoising Auto Encoders (SDAE) | 98.72 | >17小时, GPU |

L. L. C. Kasun, et al, "Representational Learning with Extreme Learning Machine for Big Data," *IEEE Intelligent Systems*, vol. 28, no. 6, pp. 31-34, 2013.

J. Tang, et al, "Extreme Learning Machine for Multilayer Perceptron," *IEEE Transactions on Neural Networks and Learning Systems*, 2015.

多层ELM应用：手语动作识别

| Methods | ELM Based | Tensor canonical correlation | Tangent bundles on special manifolds |
|------------|-----------|------------------------------|--------------------------------------|
| Accuracies | 99.4% | 85% | 93.4% |

[J. Tang, et al 2015]

Class 1:
Flat/Leftward



Class 2:
Flat/Rightward



Class 3:
Flat/Contract



Class 4:
Spread/Leftward



Class 5:
Spread/Rightward



Class 6:
Spread/Contract



Class 7:
V-shape/Leftward



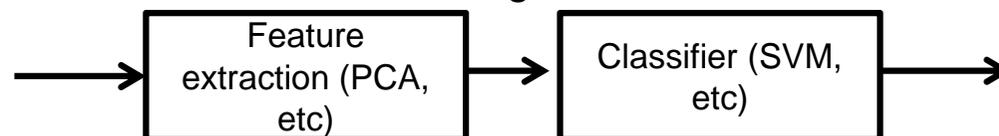
Class 8:
V-shape/Rightward



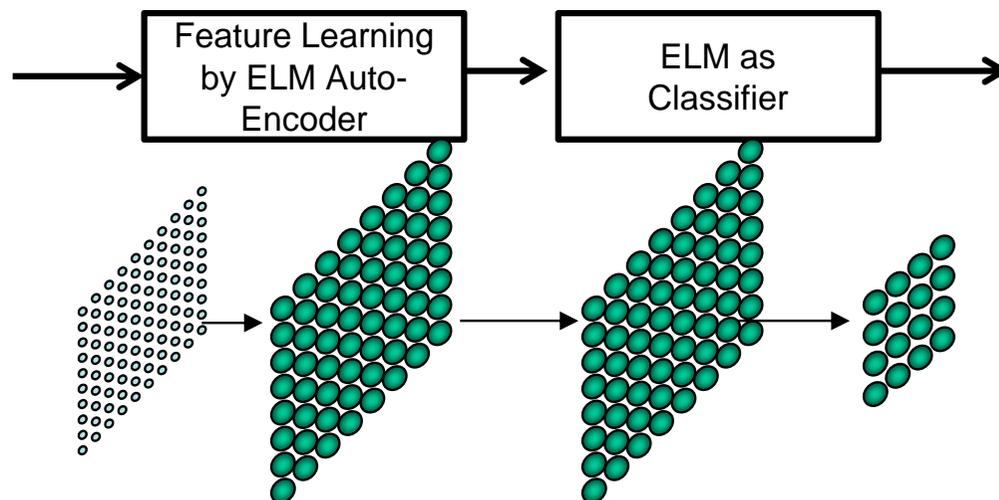
Class 9:
V-shape/Contract



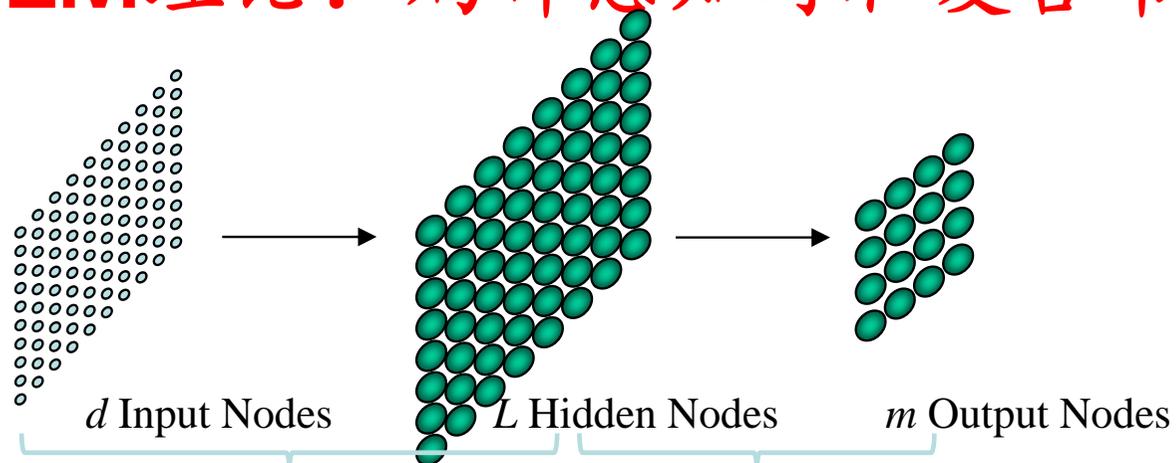
Conventional: Heterogeneous combinations



Conventional: Homogeneous combinations

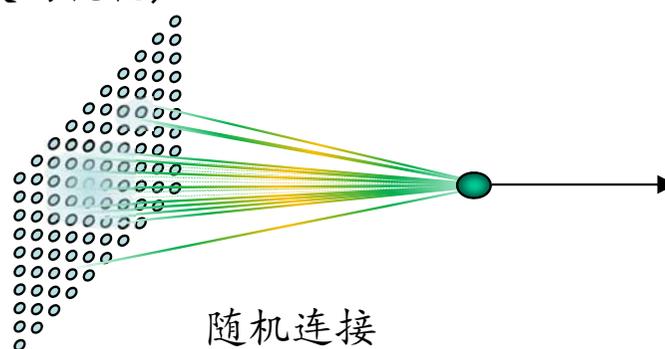
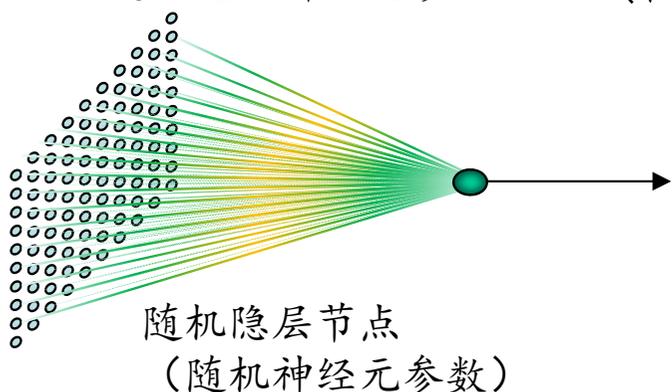


ELM理论：局部感知野和复合节点

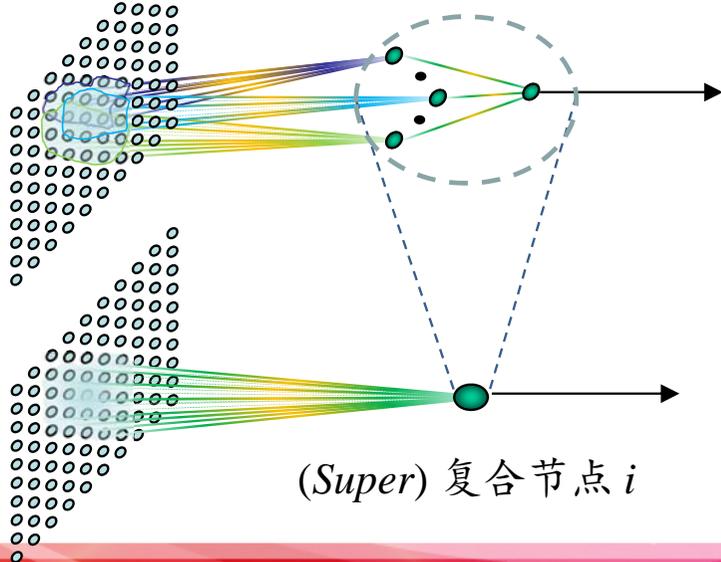
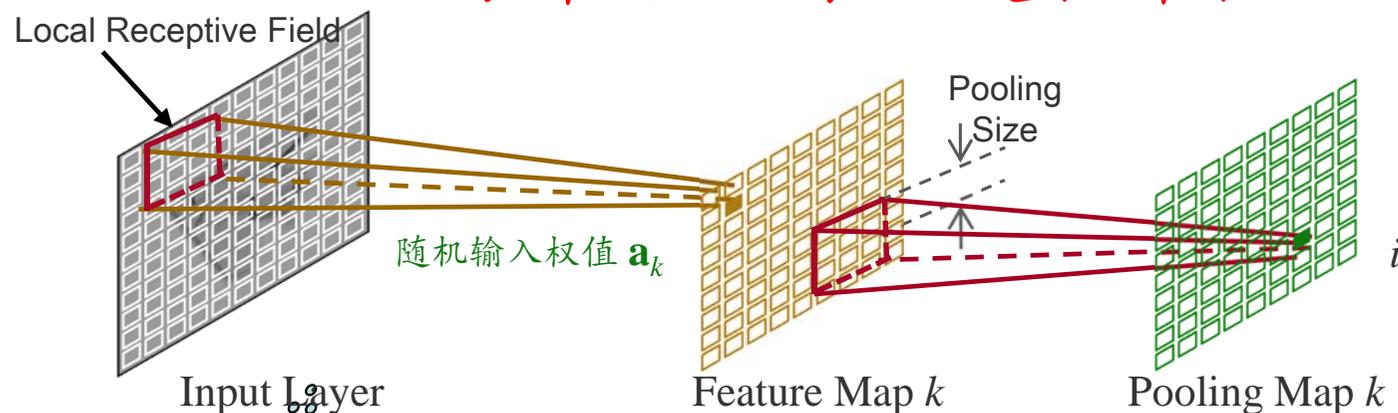


ELM特征映射
(随机隐层神经元参数)

ELM学习
(和应用相关的优化)



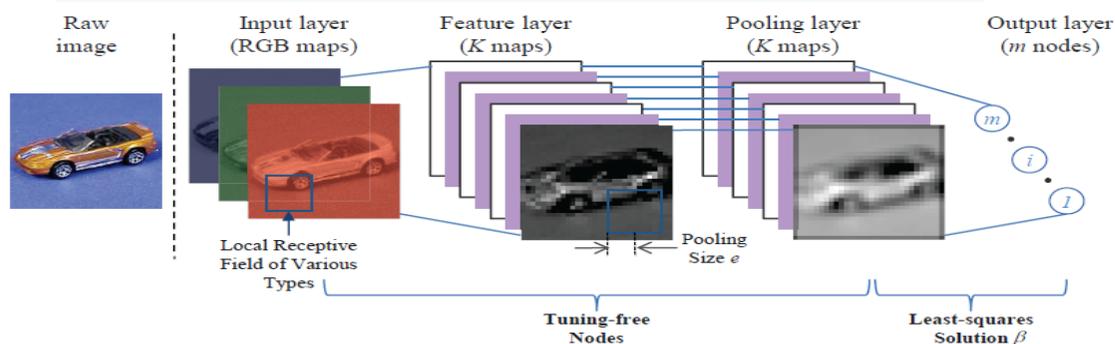
ELM理论：局部感知野和复合节点



卷积运算可以看作是一个卷积神经元（特殊的非线性阶段连续函数）。卷积神经元和Pooling可以看作是ELM局部感知野和复合节点的一个特例，但是ELM的局部感知野和复合节点可以多种。

ELM vs 深度学习

| 学习方式 | | 测试误差 |
|-------|---------------------------------|--------|
| 超限学习机 | ELM | 0.02% |
| 深度学习 | Convolutional Neural Nets (CNN) | 28.51% |
| | CNN+video (test images of COIL) | 7.75% |
| | CNN++video (COIL-like images) | 20.23% |



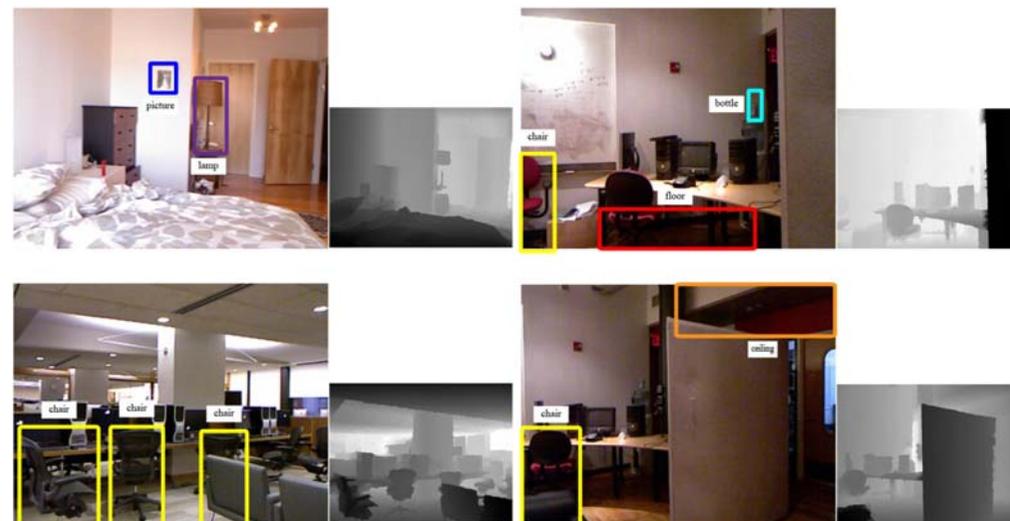
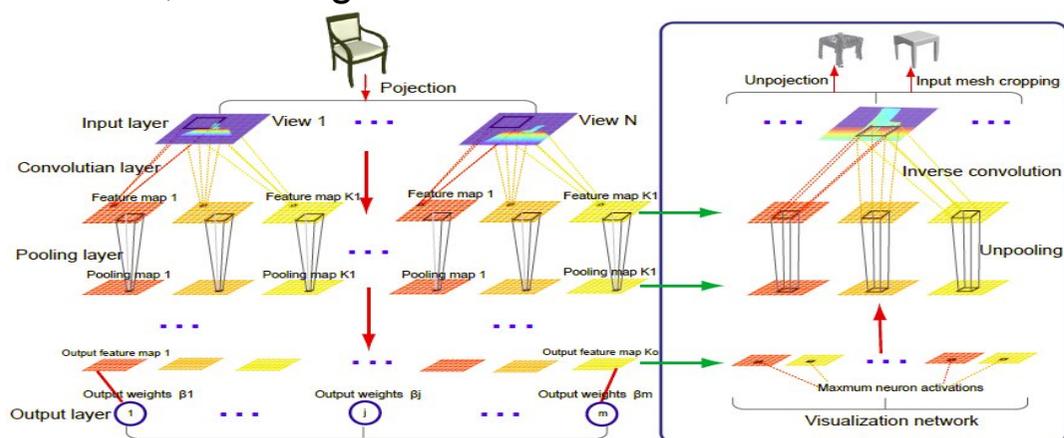
COIL Dataset: 1800 training samples, 5400 testing samples, 100 categories

Z. Bai, et al, "Generic Object Recognition with Local Receptive Fields Based Extreme Learning Machine," 2015 INNS Conference on Big Data, San Francisco, August 8-10, 2015; *Procedia Computer Science*, vol. 53, pp. 391-399, 2015

ELM vs 深度学习

| 学习方式 | 测试误差 | 训练时间 |
|---------|--------|-------------|
| MVD-ELM | 81.39% | 306.4秒(CPU) |
| CDBN | 77.32% | > 2天 (GPU) |

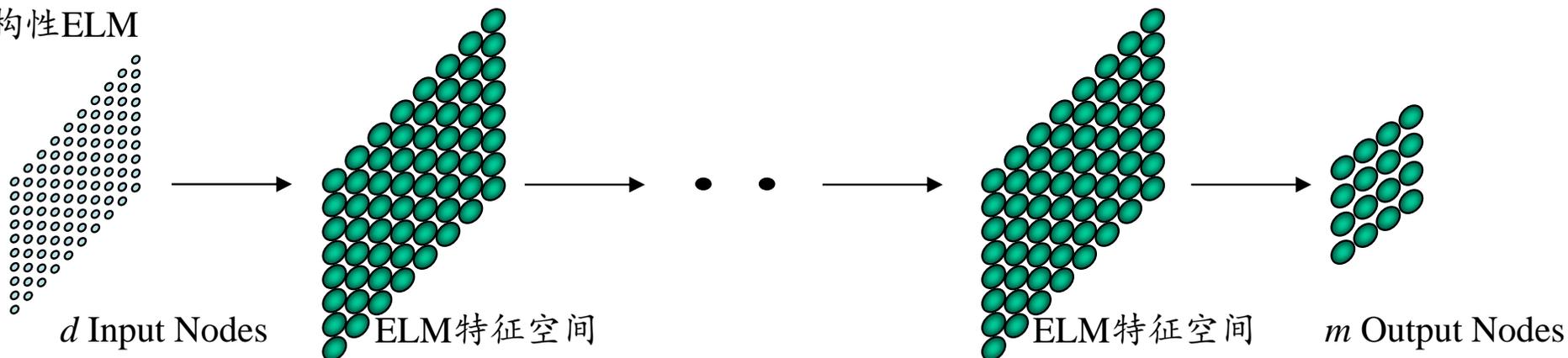
ModelNet40: 10695 training data, 8567 testing data, 40 categories



Z. Xie, et al, "Projective Feature Learning for 3D Shapes with Multi-View Depth Images," The 23rd Pacific Conference on Computer Graphics and Applications, Tsinghua University, China, October 7-9, 2015.

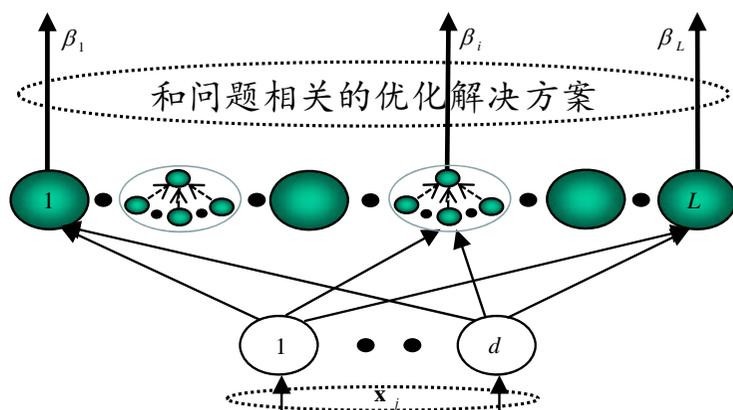
结构性 (Hierarchical) ELM

(a) 结构性ELM



Different from Deep Learning, All the hidden neurons in ELM as a whole are not required to be iteratively tuned

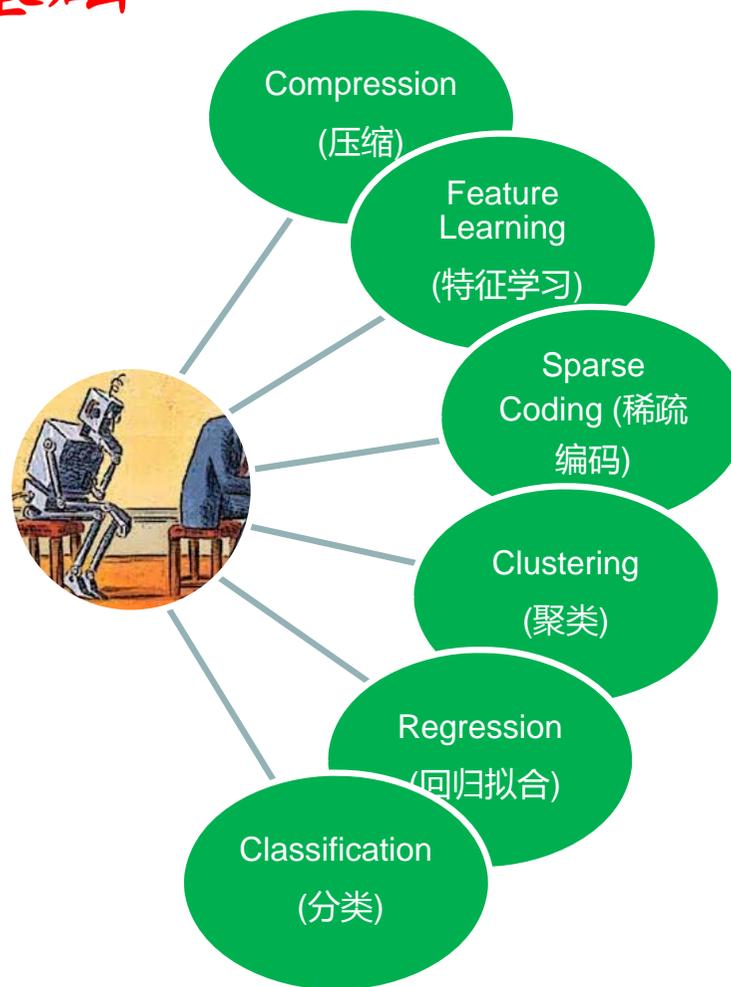
(b) ELM子网络 (包含复合隐层节点)



- 压缩
- 特征学习
- 稀疏编码
- 聚类
- 回归拟合
- 分类

G.-B. Huang, et al., "Convex Incremental Extreme Learning Machine," *Neurocomputing*, 2007.

ELM: 感知和推论的基础



ELM vs 生物学习

ELM

对网络结构大小不太敏感

对人工参数不太敏感

容易并行和硬件实现

较易微型实时学习；需要较少的时间；较易实现多通道数据实时融合和决策系统实时同步

交易实现在线串行数据学习 (sequential learning) 和数据流学习 (stream data learning)

理论上可以用小数据解决复杂的应用

易于低功耗硬件实现

学习速度极其快速，在学习精度上讲求高准确度上的动态平衡

可以先有通用算法和硬件

生物学习

对生物神经元网络模块大小不是苛求，稳定性很高（每个模块大小不等：几十~几万个神经元）

并没有什么“人”在“脑”中不时“调参”

并行和硬件实现是生物学习机制的天然实现模式

微型实时学习；天然实现多通道数据实时融合和决策系统实时同步

天然在线串行数据学习和数据流学习

用小数据解决许多复杂应用

低速的神经元并行学习

讲究高学习速度，高的准确率，但不过分“贪婪”于学习精度

总是先有普适的“脑”，再有应用



ELM vs 生物学习 vs 计算机

- 计算机之父冯·诺伊曼的困惑 [Neumann 1951, 1956]

- 和计算机需要完美硬件连接组成所不同的是，为什么一个看上去不完美的包含许多看似随机连接的（生物）神经网络却能够可靠地实现完美的学习功能？ [Rosenblatt 1958]



- 60年后...

- ELM学习理论给出回答 [Huang, et al 2006, 2007, 2008]

- 网络的整个多层结构（人工神经网络或生物网络）是结构化且有序的，但它们在某一个神经元层或神经模块片中看起来「混乱、非组织结构化」。从局部来看，「硬连线」可以是全连接或部分连接。这种全局结构化而局部随机连接的看似「不完美」结构，却正好构成了基本的完美的学习能力，包括压缩感知、特征学习、稀疏编码、聚类、回归和分类等。这就解决了冯·诺依曼对生物学习的迷惑。 [Huang 2014]
- 自然界和脑是如此的同样“美”：自然界是总体有序，局部是随机的布朗运动。 ELM学习理论：脑也是总体有序局部无序。

ELM: 目前的应用领域

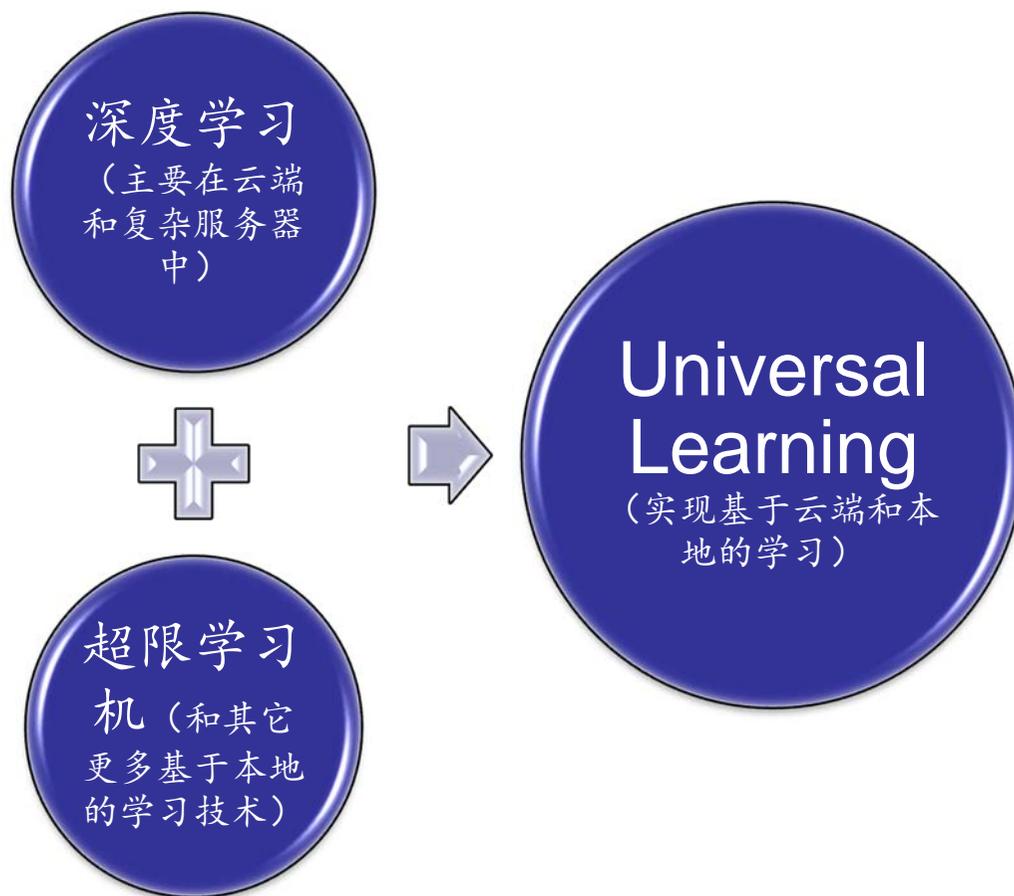


新趋势：机器学习和生物学习的交融、收敛和汇合



ELM: 做为存在于机器学习和生物学习的共同基本学习单元
(基本“学习粒子”)

新趋势：云学习和本地学习的汇合 - ELM和深度学习的互补性



再思考人工智能和机器学习的内涵和发展趋势

人工智能和机器学习是两个有重叠、相互推进又不同的范畴。

智能时代的代表不只是数据大!!

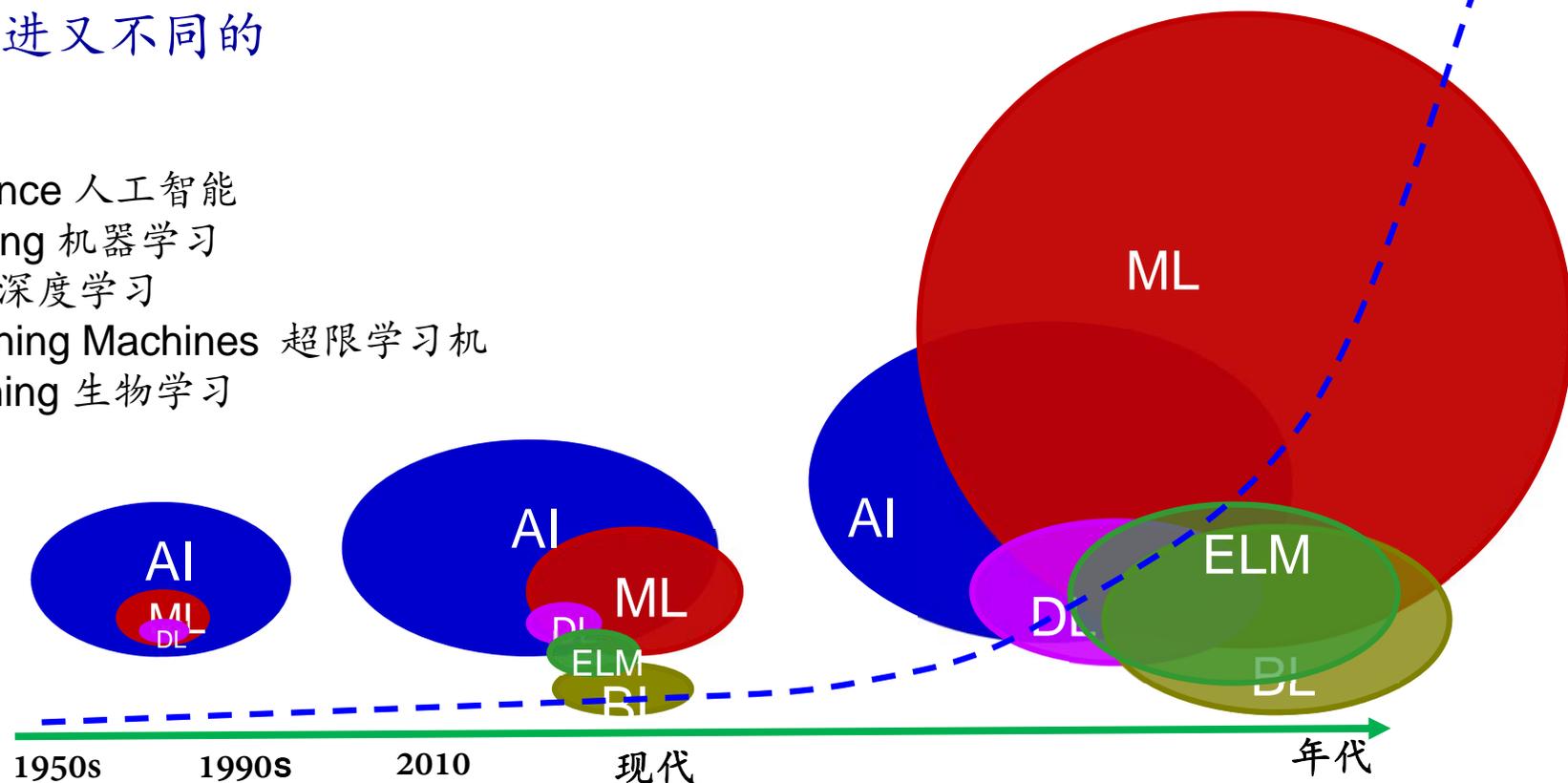
AI: Artificial Intelligence 人工智能

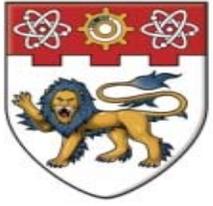
ML: Machine Learning 机器学习

DL: Deep Learning 深度学习

ELM: Extreme Learning Machines 超限学习机

BL: Biological Learning 生物学习





NANYANG
TECHNOLOGICAL
UNIVERSITY

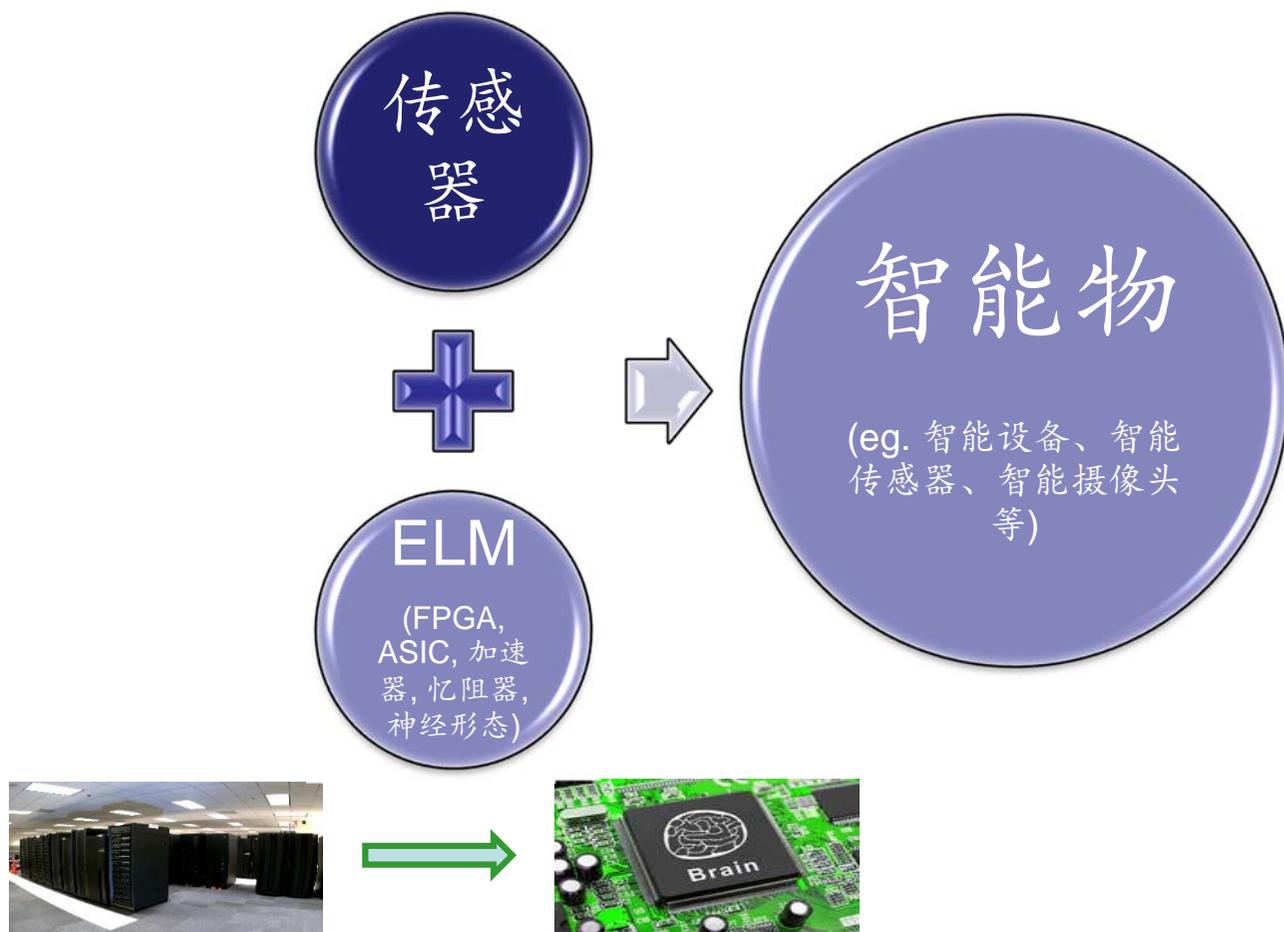
迈向

智能物联网时代的
普适学习和普适智能

Towards

Pervasive Intelligence and Pervasive
Learning in
Internet of Intelligent Things
(Not just Internet of Things)

新趋势：从物联网到智能物联网



智能物联网三阶段：物联网、智能物联网、智能物社会、 新经济模式

物联网 (Internet of Things)

- “物” (Things) 有采集数据和计算功能
- 缺乏学习能力和智能

智能物联网 (Internet of Intelligent Things)

- 基于ELM或其它嵌入智能技术的智能体
- 具有学习能力的智能材料，智能传感器，智能设备
- 移动数据中心 (比如车，列车，人，机器人等)

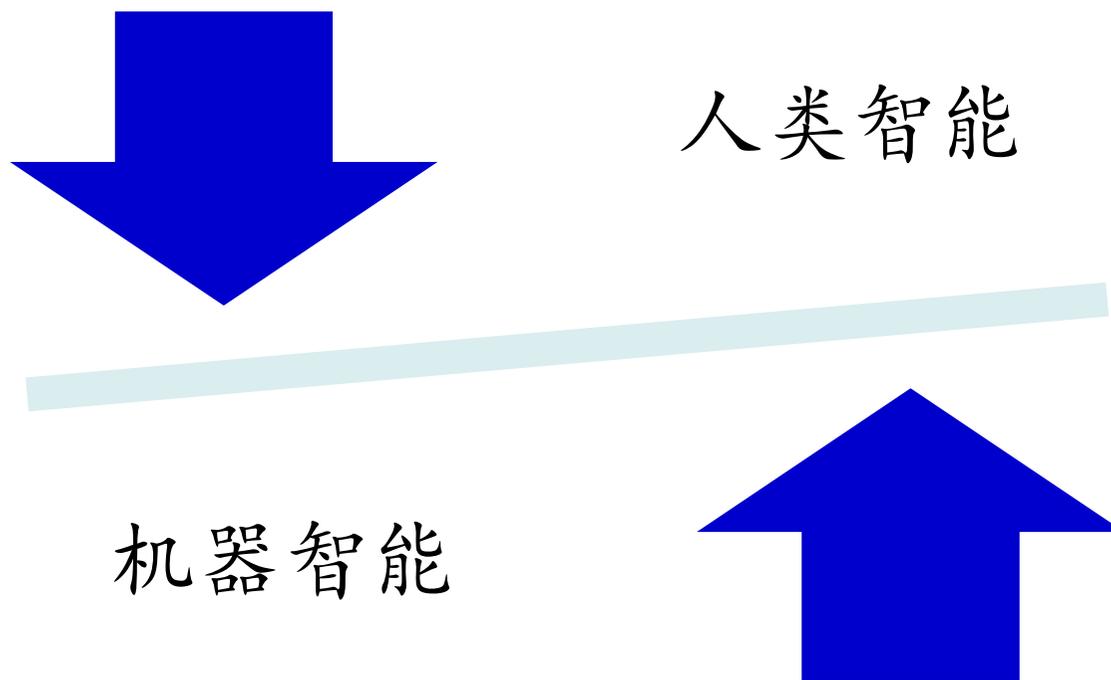
智能物社会 (Society of Intelligent Things)

- 有生命的智能和无生命的智能 (包括机器智能) 共存
- 新的经济模式出现

智能物联网三阶段：物联网、智能物联网、智能物社会



机器智能 vs 人的智能



人的能力在下降？机器智能在提高？

给定一个特定的智能任务，机器最终会超过人的能力？

智能指数 (Intelligence Index)

- 有生命和无生命的智能会普遍化
- 地球总体智能指数在迅速提升，爆发阶段即将到来

- 担心/矛盾:

- 人类和生物智能的主宰地位却受到人类创造的智能的挑战。
- 根据达尔文进化论，随着智能的普及深化，人的部分能力会退化，人也会越来越少
- 虽然非生命体智能由人类创造，却最终很可能变的不可控、不可确定性。
- 正面/负面智能？
- 最终是智能而不是智能的载体穿越时空。

