NANYANG TECHNOLOGICAL UNIVERSITY

Trend of Artificial Intelligence, Machine Learning and Data Analytics for Internet of Things

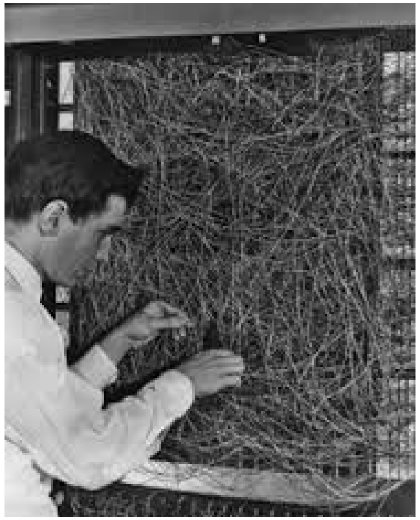
Huang Guangbin

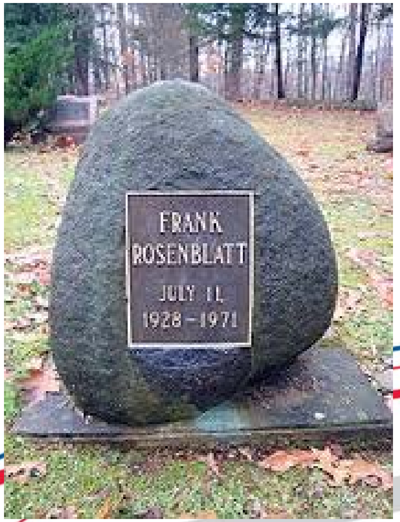
School of Electrical and Electronic Engineering

Nanyang Technological University, Singapore

Frank Rosenblatt: Perceptron

- **Cognition Dream in 60 Years Ago …**
  - "Rosenblatt made statements about the perceptron that caused a heated controversy among the fledgling AI community."
  - *Cognition*: "Based on Rosenblatt's statements, The New York Times reported the perceptron to be "the embryo of an electronic computer that [the Navy] expects will be able to **walk**, **talk**, **see**, **write**, **reproduce** itself and be **conscious** of its existence"





http://en.wikipedia.org/wiki/Perceptron

Perceptron and AI Winter

- **"AI Winter" in 1970s**
  - "Beautiful mistakes" [Minsky and Papert 1969]: Minsky claimed in his book that the simple XOR cannot be resolved by two-layer of feedforward neural networks, which "drove research away from neural networks in the 1970s, and contributed to the so-called AI winter." [Wikipedia2013]

Three Waves of Machine Learning

1950s-1980s: Warm up

Features: computers not powerful, no efficient algorithms, no enough data

Situation: Chinese people already had good dream since the inception of computers and called computers as “Electronic Brains （电脑）”

1980s-2010: Research driven

Features: computers very powerful, many efficient algorithms developed, no enough data in many cases

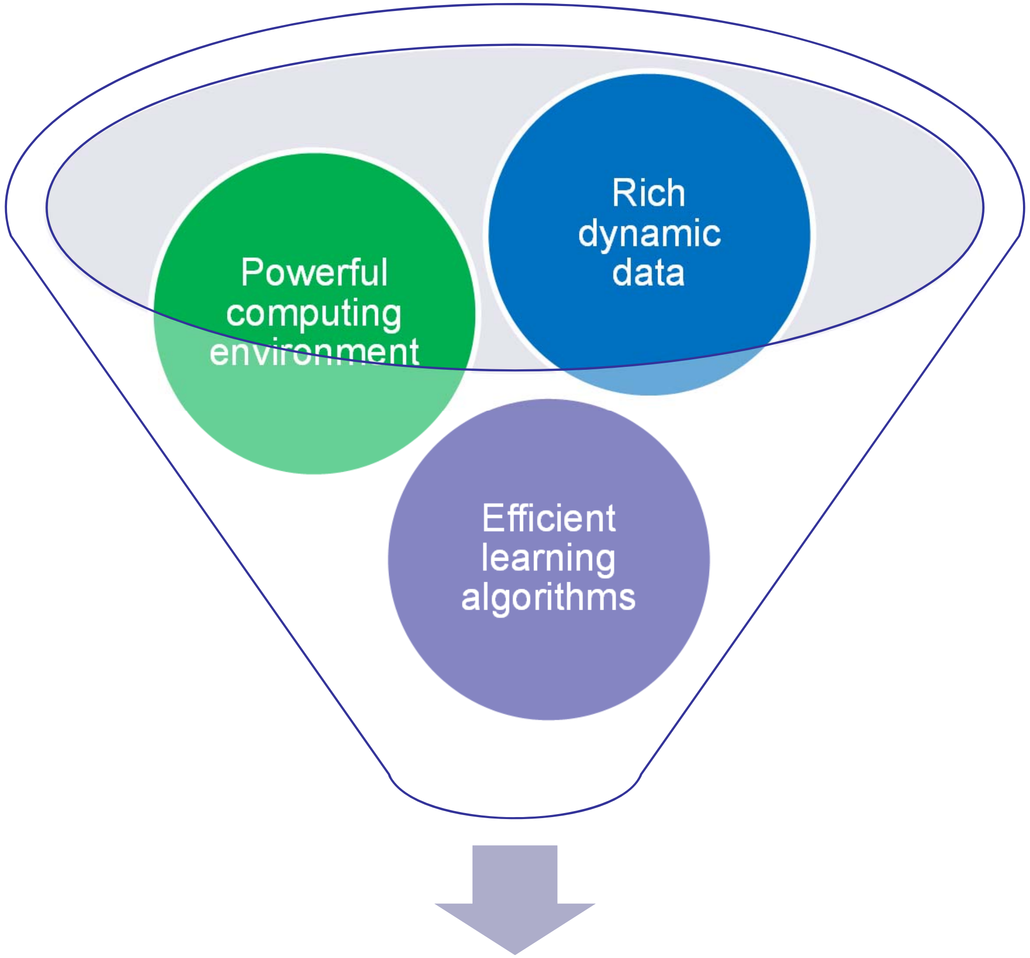
Situation: more driven by researchers instead of industries

2010 – Present: Data driven

Features: computers powerful enough, powerful and smart computing sensors/devices everywhere, huge data coming. Efficient algorithms under way

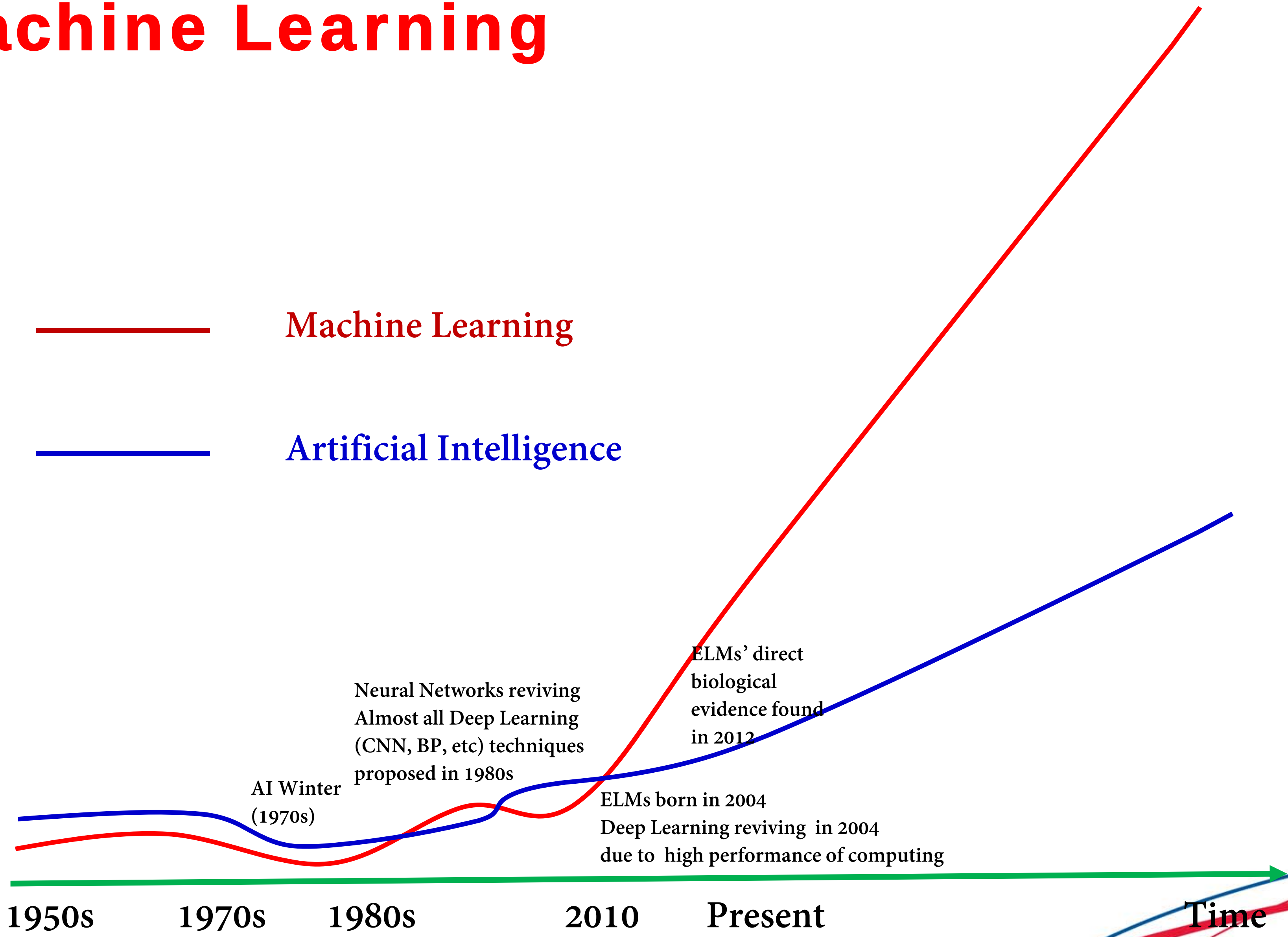
Situation; No matter you admit or not, we have to rely on machine learning from now on

Necessary Conditions of Machine Learning Era

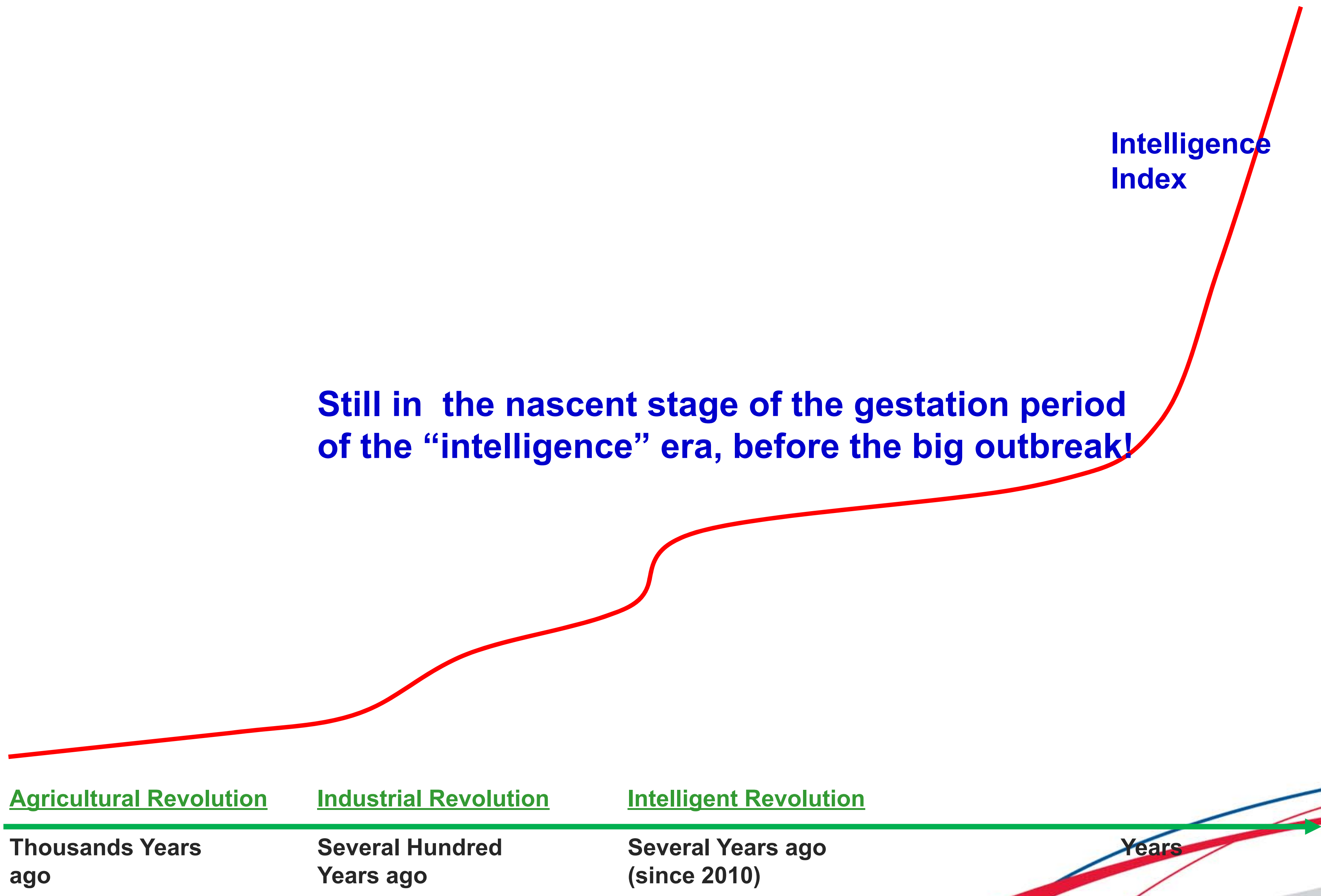


Three necessary conditions of true machine learning era, which have been fulfilling since 2010

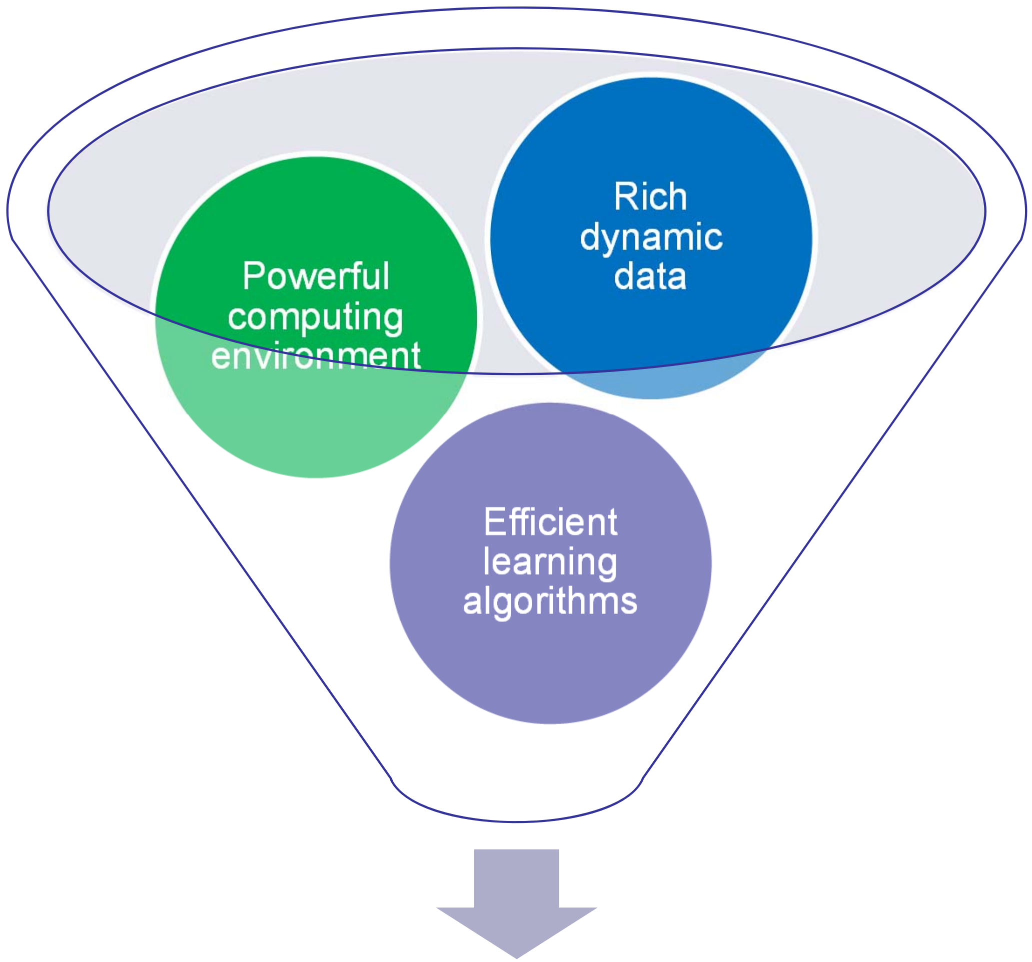
Rethink Artificial Intelligence and Machine Learning



3rd Revolution in Productivity History



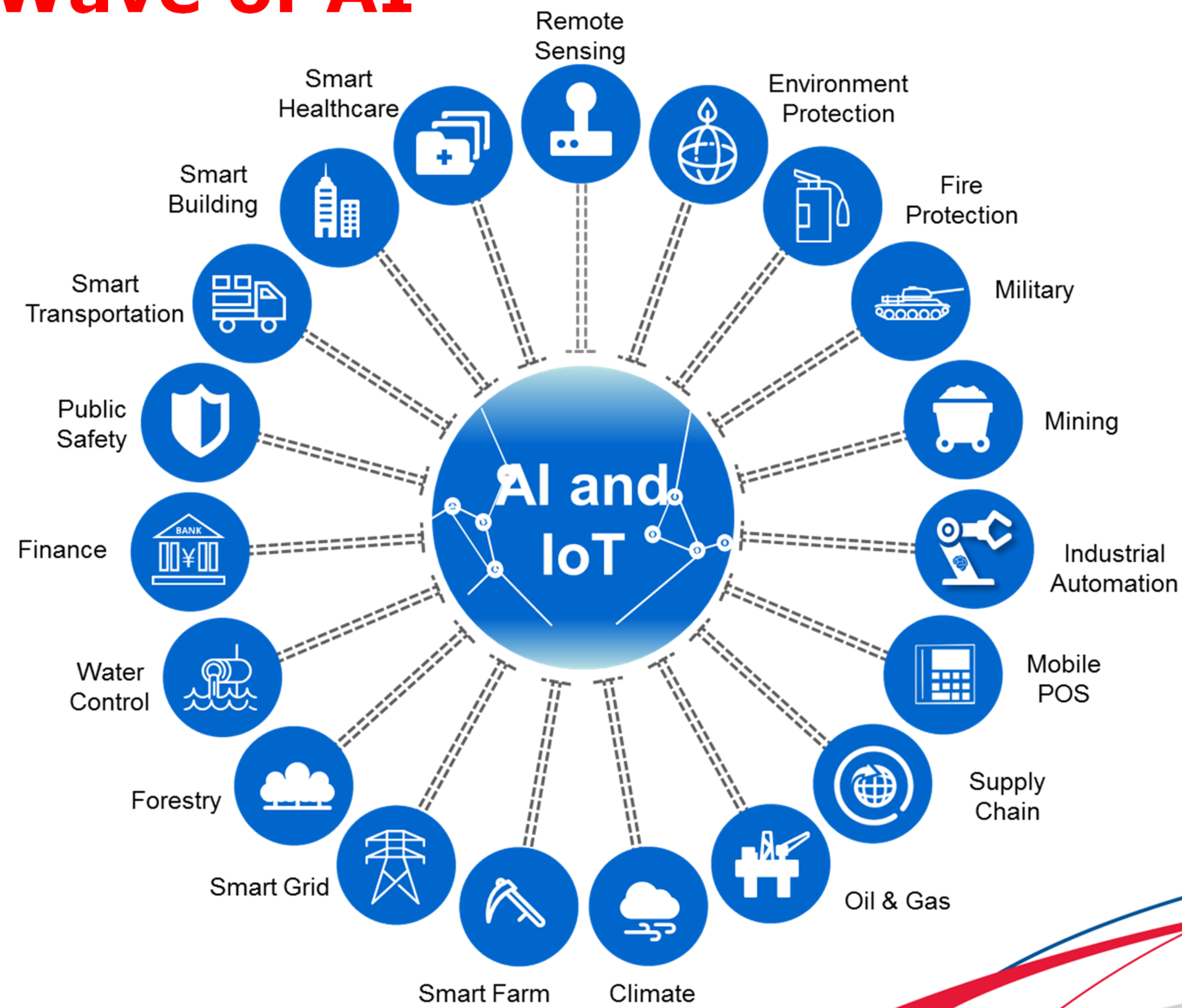
Necessary Conditions of Machine Learning Era



Three necessary conditions of true machine learning era, which have been fulfilling since 2010

Intelligence: from Cloud to Local

Synergy of AI and Internet of Things: Next Wave of AI



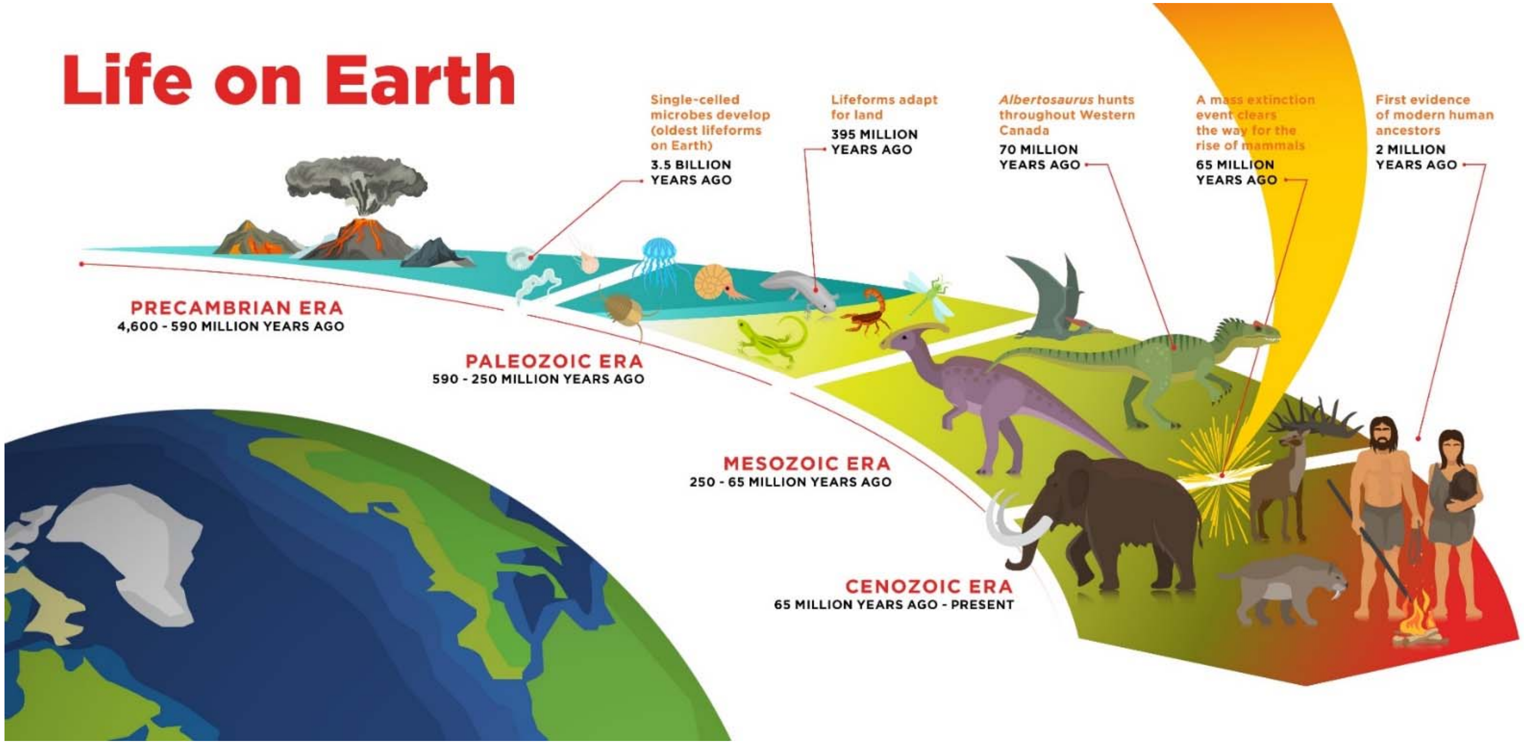
Artificial Learning and Biological Learning: Contradictory Approaches

Conventional Learning Methods	Biological Learning
Very sensitive to network size	Stable in a wide range (tens to thousands of neurons in each module)
Difficult for parallel implementation	Parallel implementation
Difficult for hardware implementation	"Biological" implementation
Very sensitive to user specified parameters	Free of user specified parameters
Different network types for different type of applications	One module possibly for several types of applications
Time consuming in each learning point	Fast in micro learning point
Difficult for online sequential learning	Nature in online sequential learning
"Greedy" in best accuracy	Fast speed and high accuracy
Big data for simple applications	Small data for complicated applications
"Brains (devised by conventional learning methods)" are chosen after applications are present	Brains are built before applications

ELM: New Theories

for Machine Learning and Biological Learning

Essential Learning Algorithms in Brains

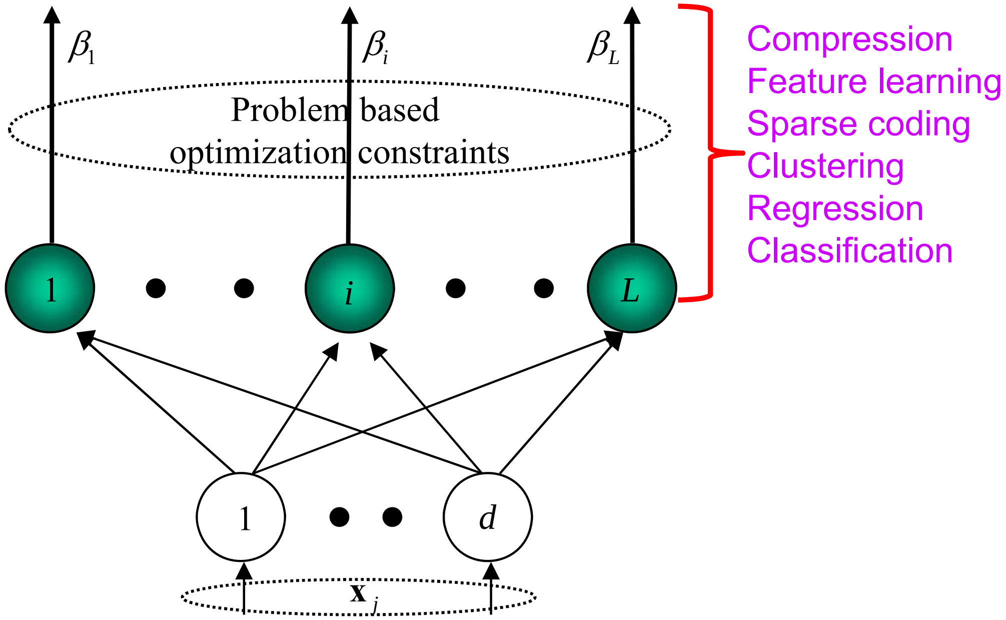


Our question: Do millions billions of human brains have common basic learning algorithms (e.g. atomic learning elements) in the past millions of years since modern human ancestors or even earlier (if including other living things)?

ELM answer: The core basic learning elements should be independent of particular applications. This can address:

1) **The reasons why brains with ultra low frequency of neurons can have faster learning and reasoning capabilities**
2) **The resilience of living brains**

Basic ELM



Almost any nonlinear piecewise continuous hidden nodes: $h_i(\mathbf{x}) = G_i(\mathbf{a}_i, b_i, \mathbf{x})$, including sigmoid networks, RBF networks, threshold networks, fuzzy inference systems, fully complex networks, high-order networks, wavelet networks, Gabor filters, *convolutional neural networks*, etc.

Output function of "generalized" SLFNs:

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

The hidden layer output function (hidden layer mapping, ELM feature space):

$$\boldsymbol{h}(x) = [G(\mathbf{a}_1, b_1, \mathbf{x}), \cdots, G(\mathbf{a}_L, b_L, \boldsymbol{x})]$$

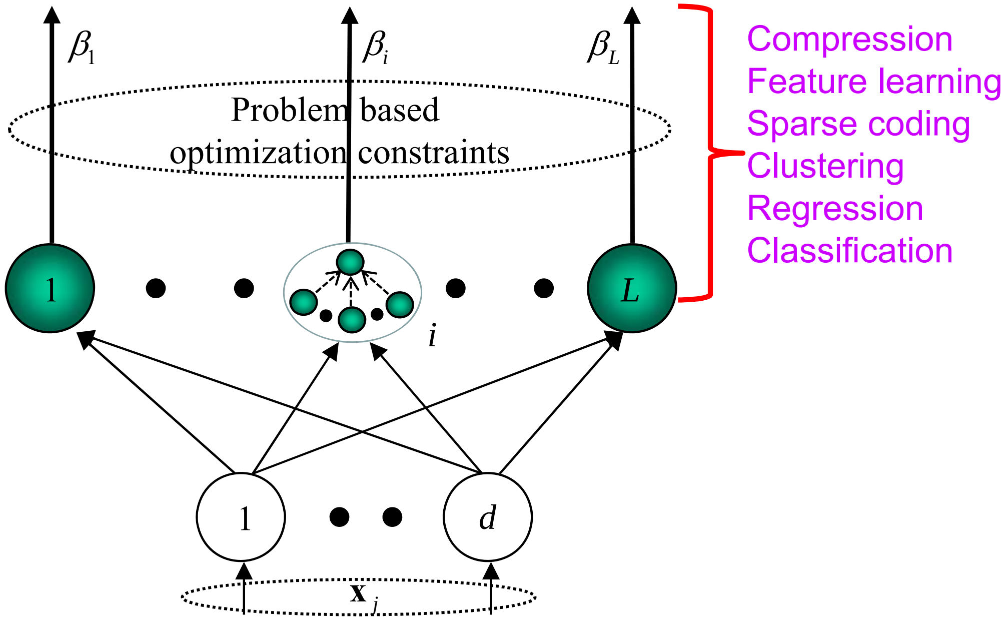
The output functions of hidden nodes can be but not limited to:

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

ELM: Super Hidden Nodes



A hidden node can be a small network of many neurons

Although we don't know ***biological neurons***' true output functions, most of them are nonlinear piecewise continuous and covered by ELM theories.

Output function of "generalized" SLFNs:

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

The hidden layer output function (hidden layer mapping, ELM feature space):

$$\boldsymbol{h}(x) = [G(\mathbf{a}_1, b_1, \mathbf{x}), \cdots, G(\mathbf{a}_L, b_L, \boldsymbol{x})]$$

The output functions of hidden nodes can be but not limited to:

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

G.-B. Huang and L. Chen, "Convex Incremental Extreme Learning Machine," *Neurocomputing*, vol. 70, pp. 3056-3062, 2007.

ELM Theory

- **New Learning Theory** - *Learning can be made without iteratively tuning (artificial) hidden nodes (or hundred types of biological neurons)* even though the modeling of biological neurons may be *unknown* as long as they are nonlinear piecewise continuous: [Huang, et al 2006, 2007, 2014, 2015] :
  - Unlike conventional learning theories and tenets, we **first time *seriously*** asked: "Does a (*biological*) learning system really need to tune hidden neurons?"
  - ELM theories point out General (*biological*) Learning systems: *locally* disordered (without being tuned) but *globally* structured, brains are as beautiful as physical world which is *globally* structured but *locally* disordered (e.g., *Brownian motions*).
  - Different from conventional learning tenets, ELM theory: As long as (*artificial* or *biological*) neural learning systems have (universal) learning capabilities, neurons need not be tuned.
    - Inherited from ancestors, transferred from one system to another, randomly generated

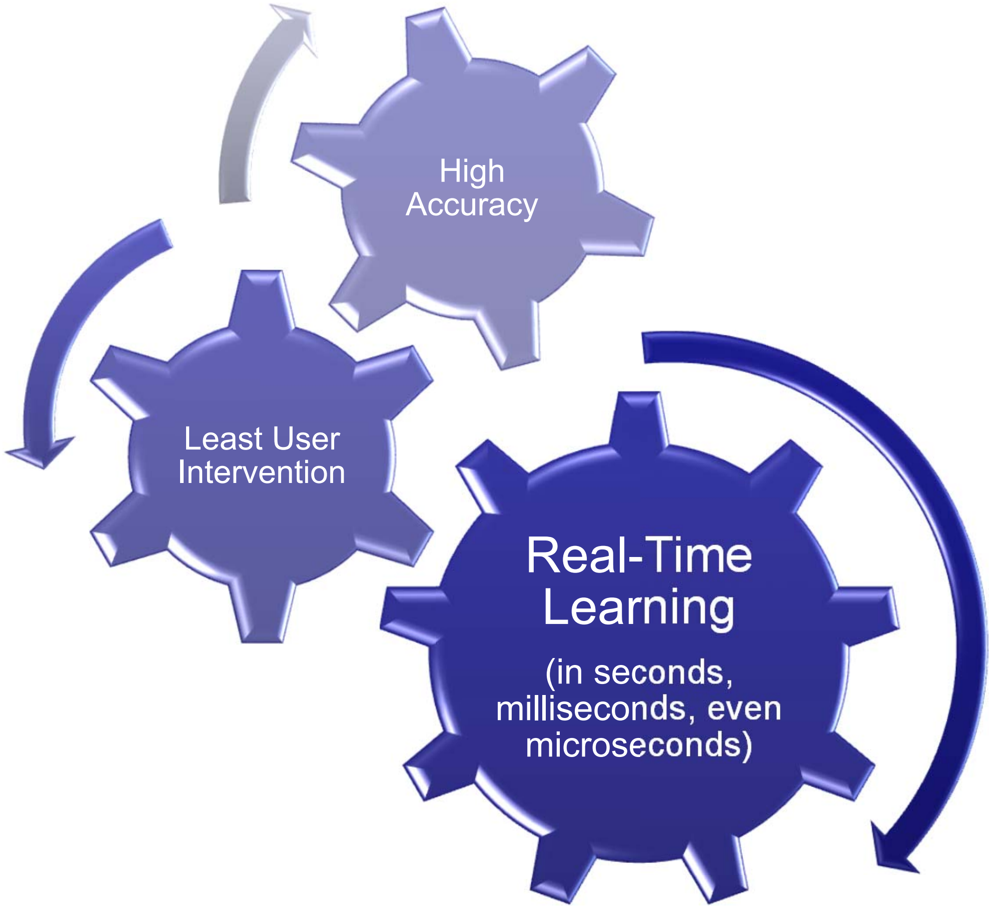
G.-B. Huang, et al., "Universal approximation using incremental constructive feedforward networks with random hidden nodes," *IEEE Transactions on Neural Networks*, vol. 17, no. 4, pp. 879-892, 2006.
G.-B. Huang and L. Chen, "Convex incremental extreme learning machine," *Neurocomputing*, vol. 70, 2007.
G.-B. Huang, "An insight into extreme learning machines: random neurons, random features and kernels," *Cognitive Computation*, vol. 6, 2014.
G.-B. Huang, "What are extreme learning machines? Filling the gap between frank Rosenblatt's dream and John von Neumann's puzzle," *Cognitive Computation*, vol. 7, 2015.

ELM Theory

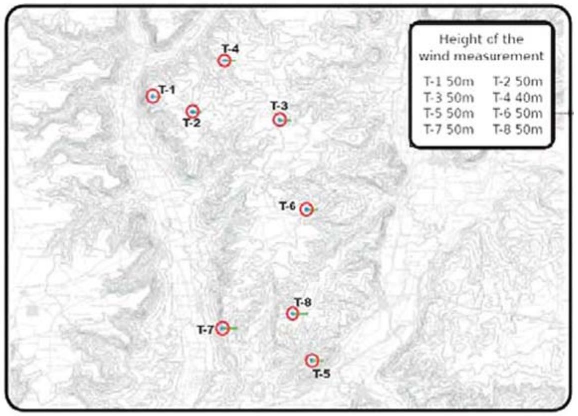
- **New Learning Theory** - *Learning Without Iteratively Tuning Hidden Neurons in general architectures*: Networks with any nonconstant piecewise continuous hidden neurons can **approximate** any continuous target function with any small error and can also **separate** any disjoint regions without tuning hidden neurons. [Huang, et al 2004, 2006, 2007]
  - All these hidden node parameters can be randomly generated (according to any continuous distribution including but not limited to uniform distribution, Gaussian distribution, etc) without training data. To inherit neurons from their ancestors is another option.
  - That is, for any continuous target function $f(\mathbf{x})$ and any randomly generated sequence $\{(\mathbf{a}_i, b_i)_{i=1}^{L}\}$, $\lim_{L\to\infty}\|f(\mathbf{x}) - f_L(\mathbf{x})\| = \lim_{L\to\infty}\left\|f(\mathbf{x}) - \sum_{i=1}^{L} \boldsymbol{\beta}_i G(\mathbf{a}_i, b_i, \mathbf{x})\right\| = 0$ holds with probability one if $\boldsymbol{\beta}_i$ is chosen to minimize $\|f(\mathbf{x}) - f_L(\mathbf{x})\|$, $\forall i$. [Huang, et al 2006, 2007]

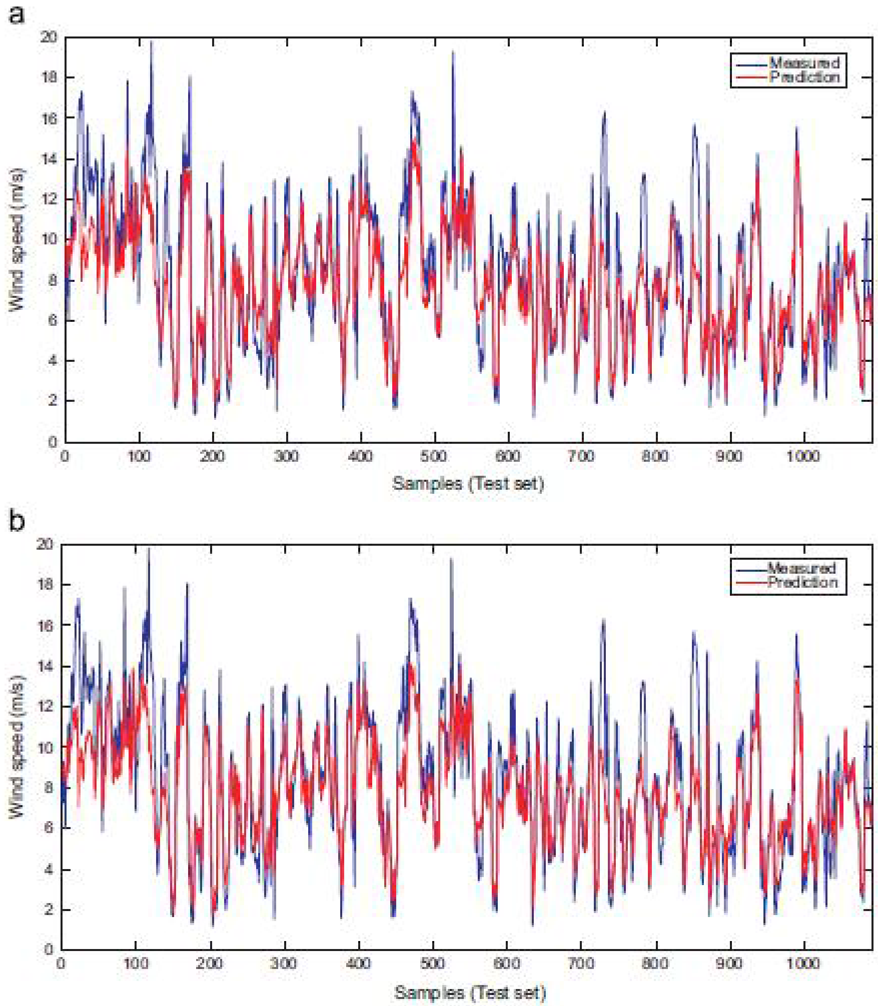
G.-B. Huang, et al., "Universal approximation using incremental constructive feedforward networks with random hidden nodes," *IEEE Transactions on Neural Networks*, vol. 17, no. 4, pp. 879-892, 2006.
G.-B. Huang and L. Chen, "Convex incremental extreme learning machine," *Neurocomputing*, vol. 70, pp. 3056-3062, 2007.
G.-B. Huang, et al, "Extreme learning machine for regression and multiclass classification," IEEE Transactions on Systems, Man, and Cybernetics - Part B: Cybernetics, vol. 42, no. 2, pp. 513-529, 2012.

Essential Considerations of ELM



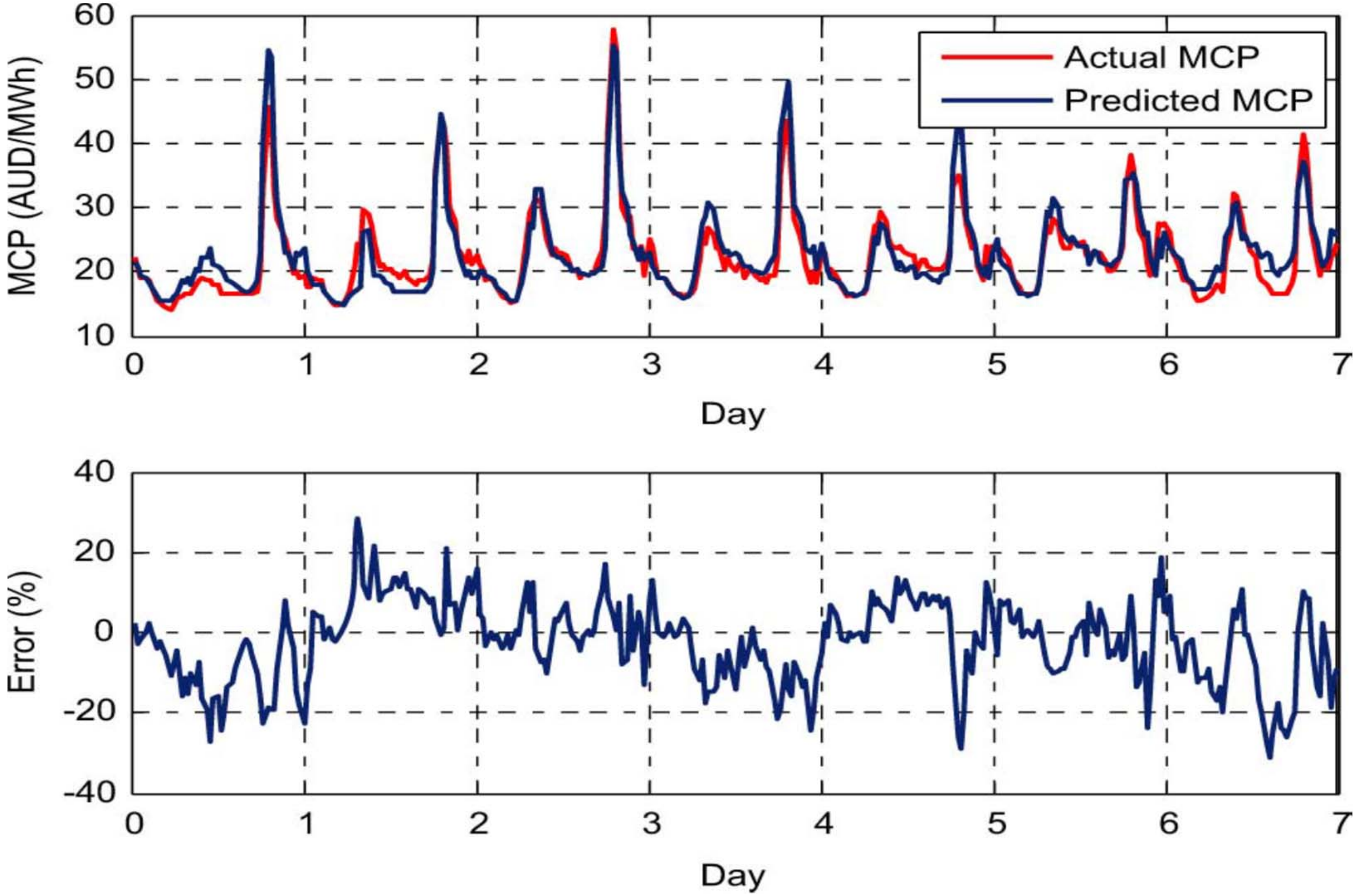
Real Operation of Wind Farms





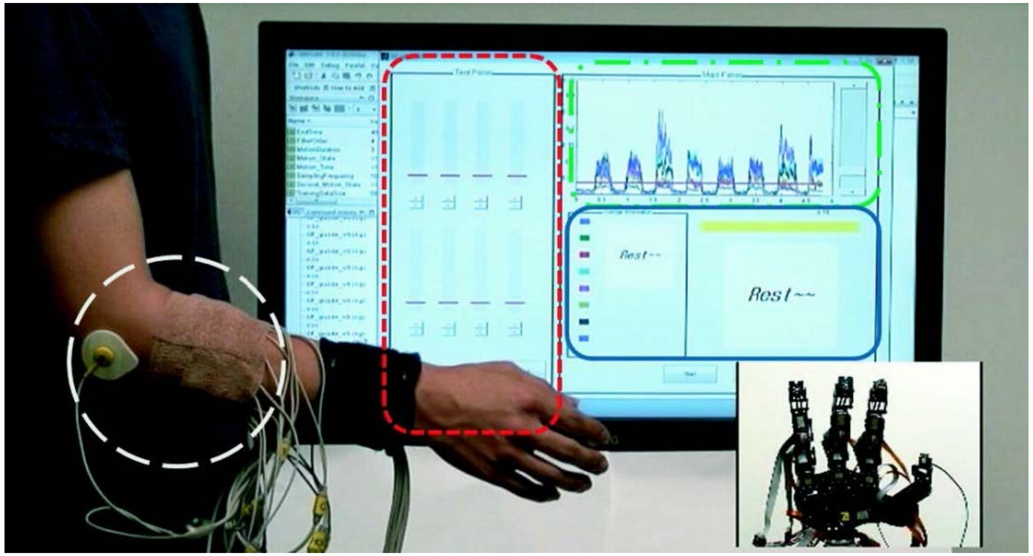
Situation of the wind measuring towers in Spain and within the eight wind farms. Wind speed prediction in tower 6 of the considered wind farm in Spain obtained by the ELM network (prediction using data from 7 towers). (a) Best prediction obtained and (b) worst prediction obtained. [Saavedra-Moreno, et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

Electricity Price Forecasting

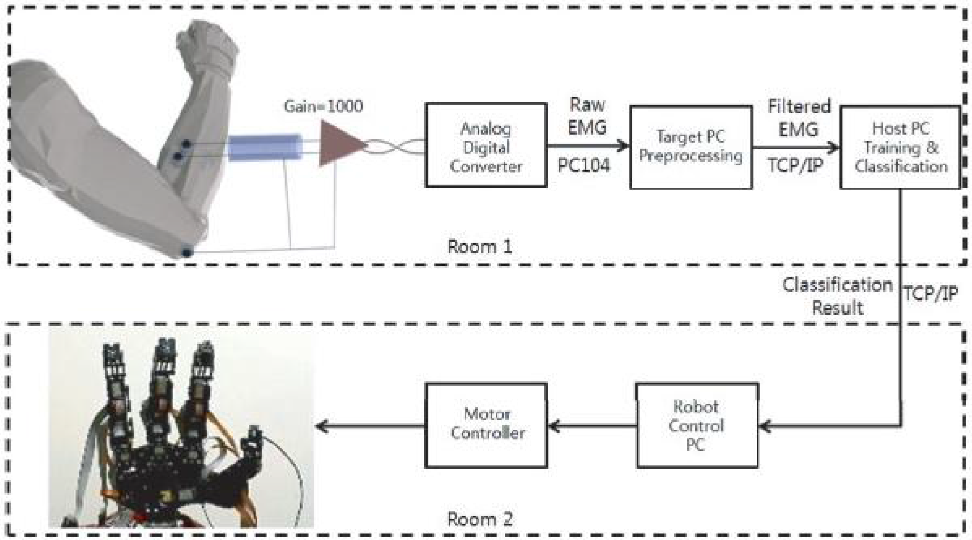


Average results of market clearing prices (MCP) forecast by ELM in winter: Trading in the Australian national electricity market (NEM) is based on a 30-min trading interval. Generators submit their offers every 5 min each day. Dispatch price is determined every 5 min and 6 dispatch prices are averaged every half-hour to determine the regional MCPs. In order to assist decision-making process for generators, there are totally 48 MCPs needed to be predicted at the same time for the coming trading day. [Chen, et al, 2012]

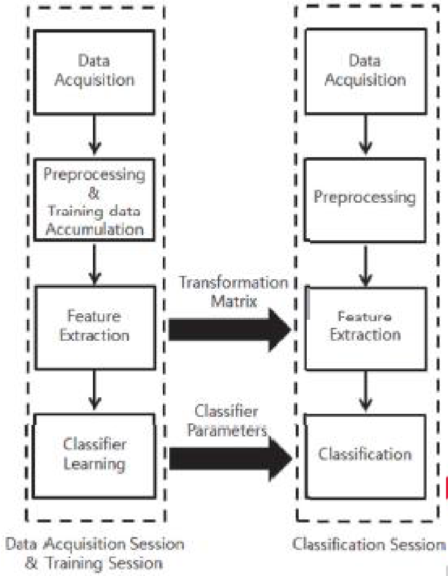
Remote Control of a Robotic Hand



- An eight wrist motions offline classification using linear support vector machines with little training time (under 10 minutes).
- This study shows human could control the remote side robot hand in real-time using his or her sEMG signals with less than 50 seconds recorded training data with ELM.[Lee, et al 2011]



(a)



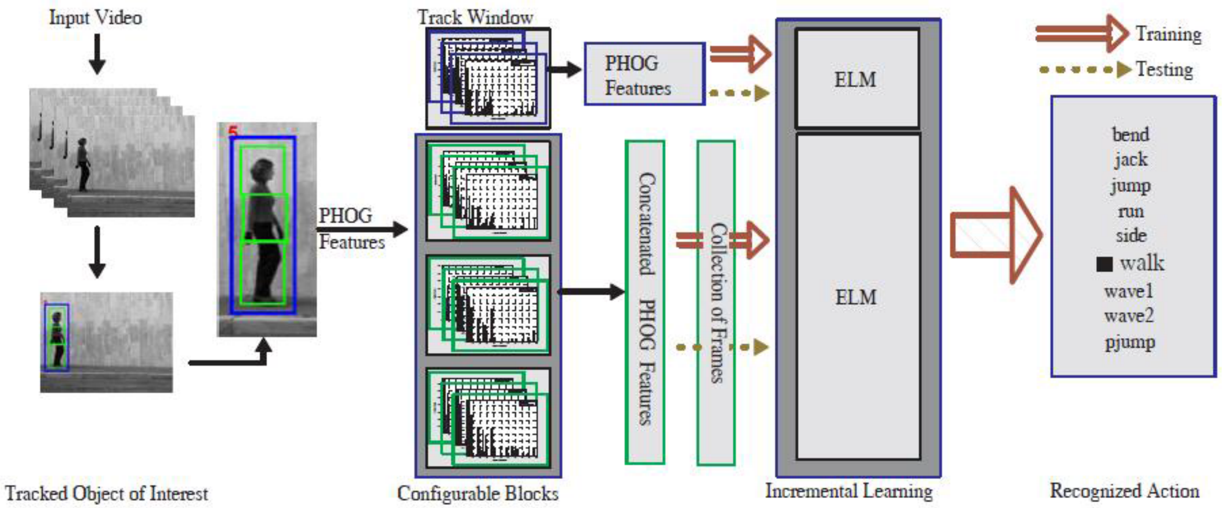
(b)

Human Action Recognition

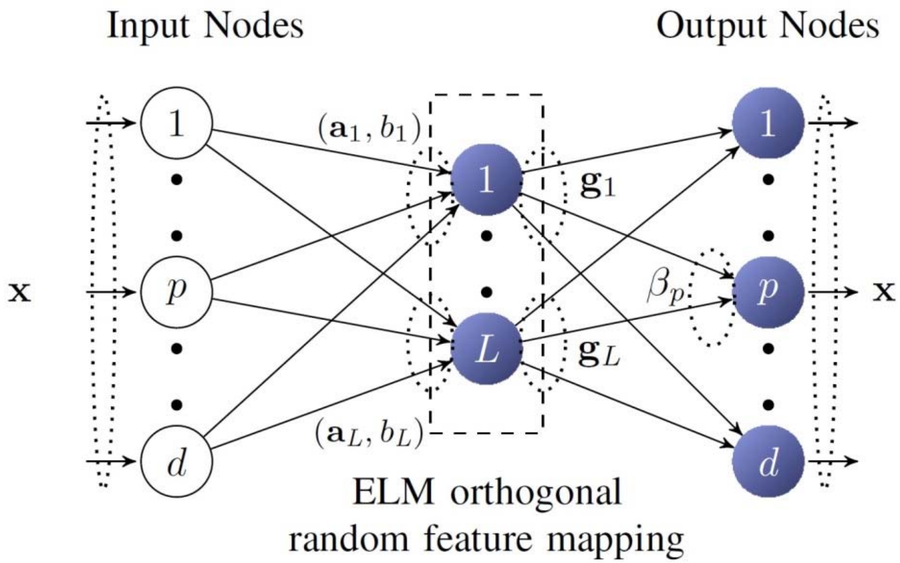
Weizmann dataset											
Methods	OS-ELM Based				[2]	[32]	[14]	[36]	[41]	[30]	[11]
Frames	1/1	3/3	6/6	10/10	-	-	-	-	-	-	-
Accuracy	100.0	100.0	100.0	100.0	100.0	72.8	98.8	100.0	97.8	99.44	100.0

KTH dataset											
Methods	OS-ELM Based				[14]	[36]	[30]	[21]	[27]	[9]	[44]
Frames	1/1	3/3	6/6	10/10	-	-	-	-	-	-	-
Accuracy	92.8	93.5	95.7	96.1	91.7	92.7	94.83	95.77	97.0	96.7	95.7

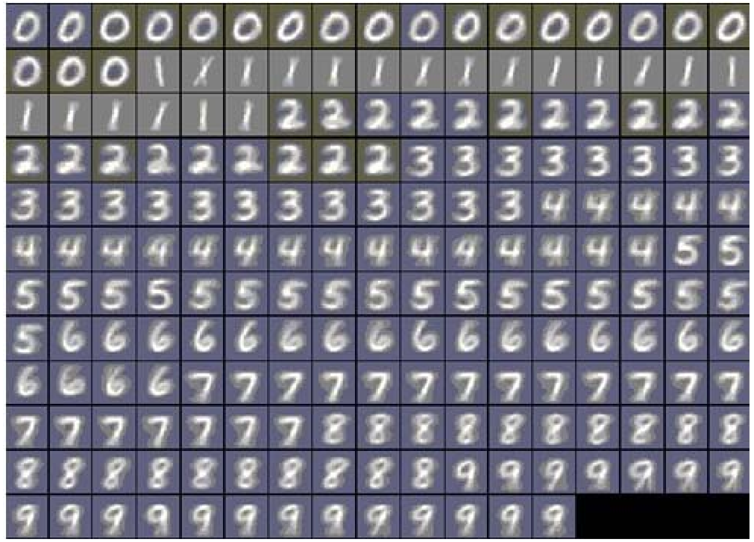
[Minhas, et al 2012]



ELM as Auto-Encoder (ELM-AE)

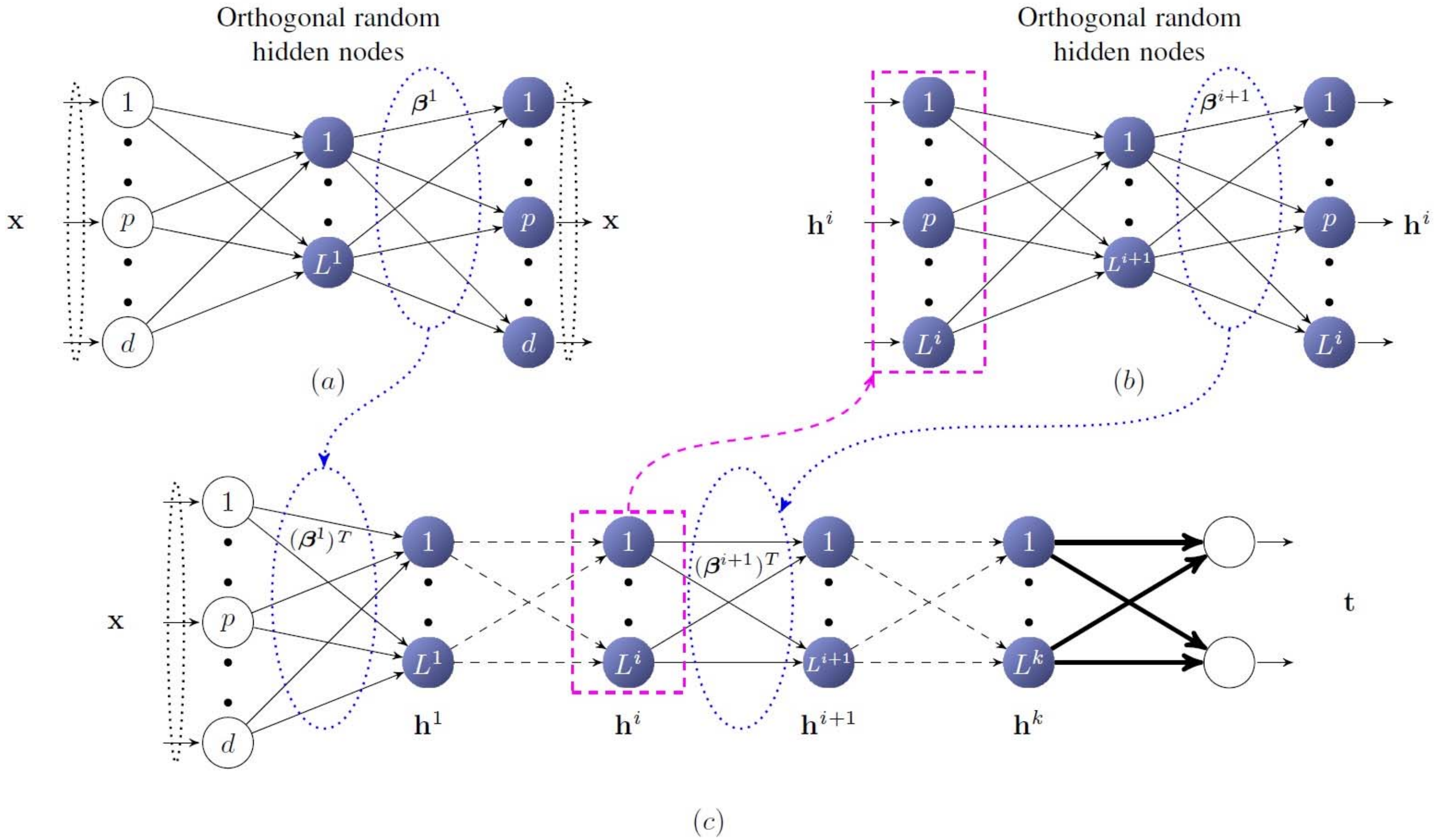


$d > L$: Compressed Representation
$d = L$: Equal Dimension Representation
$d < L$: Sparse Representation



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

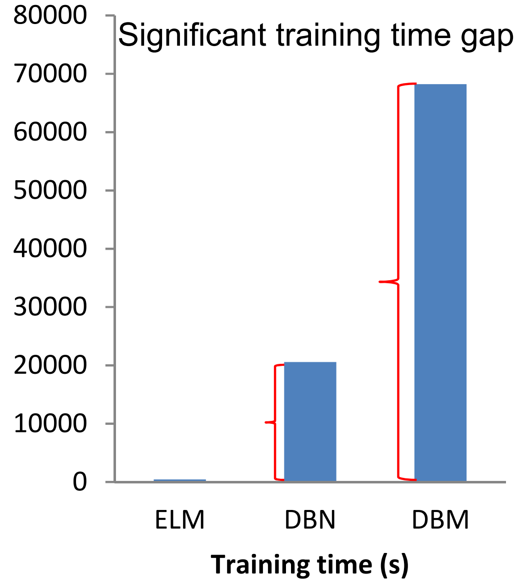
ELM as Auto-Encoder (ELM-AE)



ELM-AE based multi-Layer ELM (ML-ELM): Different from Deep Learning, no iteration is required in tuning the entire multi-layer feedforward networks

MNIST OCR

Learning Methods	Testing Accuracy	Training Time
H-ELM [Unpublished]	**~99.6%**	**Minutes (CPU)**
H-ELM [J. Tang, et al, 2015]	**99.14**	**281.37s (CPU)**
Multi-Layer ELM (784-700-700-15000-10) [Huang, et al 2013]	**99.03 ± 0.04**	**444.7s (CPU)**
Deep Belief Networks (DBN) (748-500-500-2000-10)	98.87	20580s **(5.7 hours, GPU)**
Deep Boltzmann Machines (DBM) (784-500-1000-10)	99.05	68246s **(19 hours, GPU)**
Stacked Auto Encoders (SAE)	98.6	**> 17 hours, GPU**
Stacked Denoising Auto Encoders (SDAE)	98.72	**> 17 hours, GPU**
[Huang, et al 2013]		



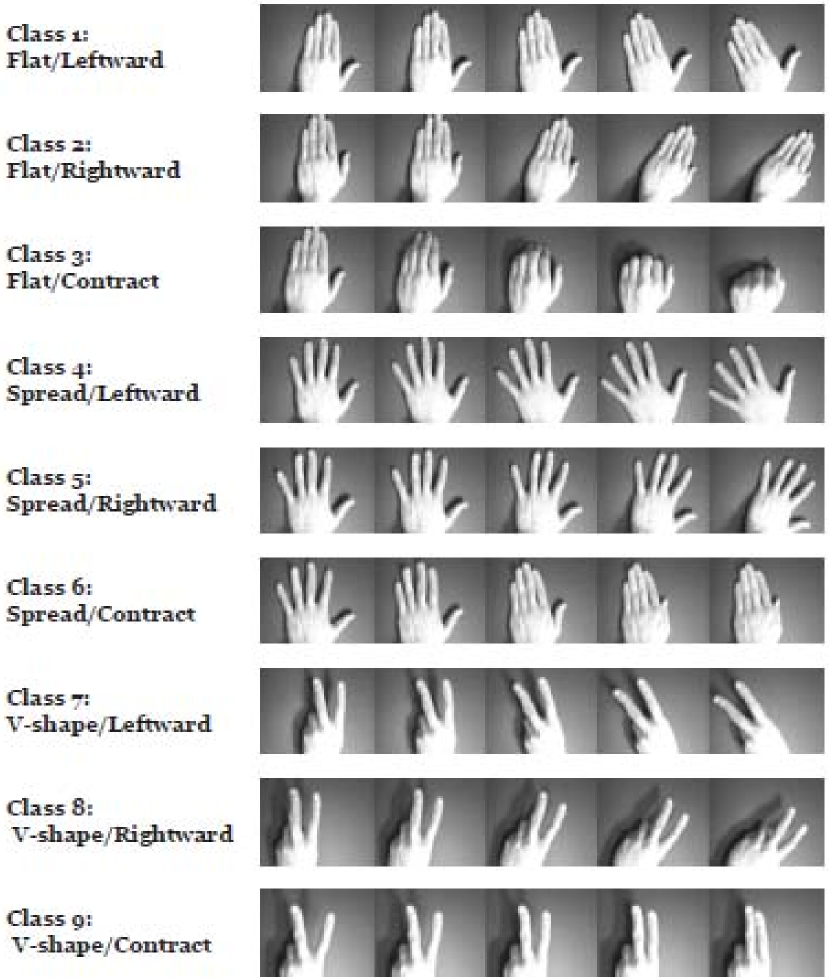
L. L. C. Kasun, et al, "Representational Learning with Extreme Learning Machine for Big Data," *IEEE Intelligent Systems*, vol. 28, no. 6, pp. 31-34, 2013.

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

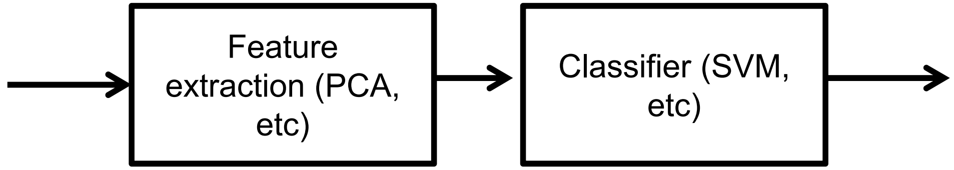
Human Action Recognition

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

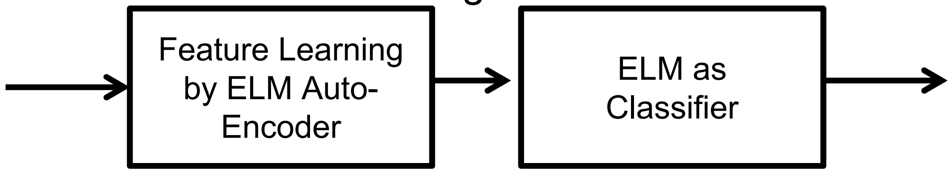
[J. Tang, et al 2015]

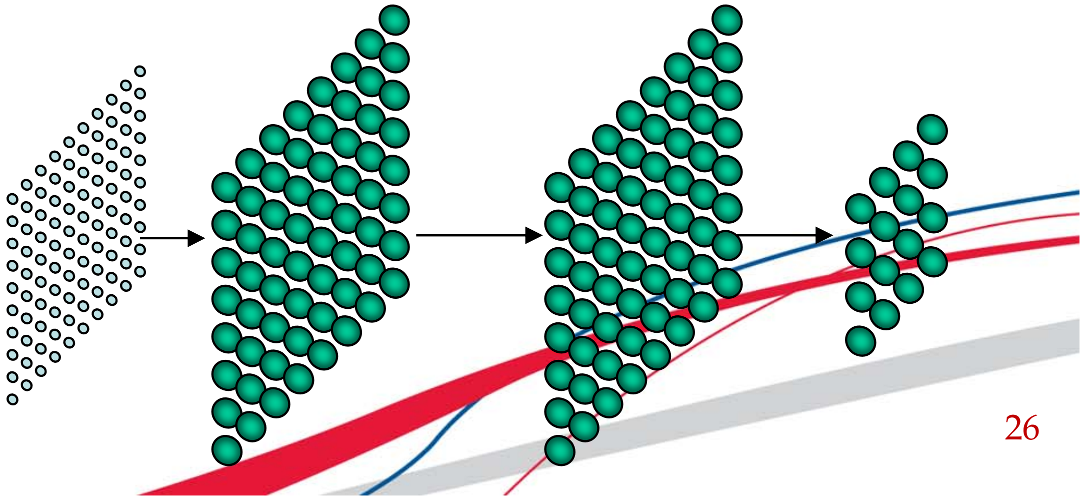


Conventional: Heterogeneous combinations



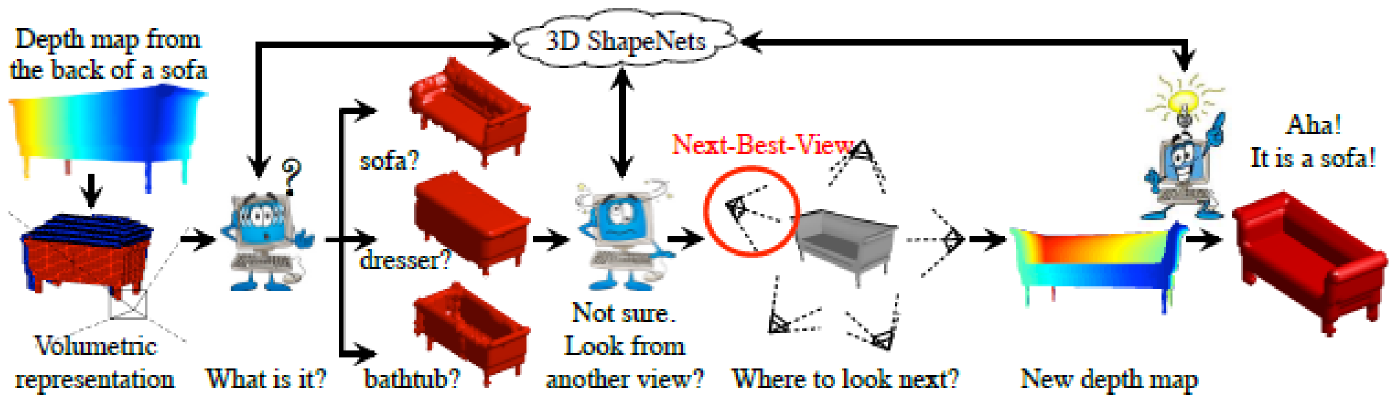
Conventional: Homogeneous combinations





ELM vs Deep Learning

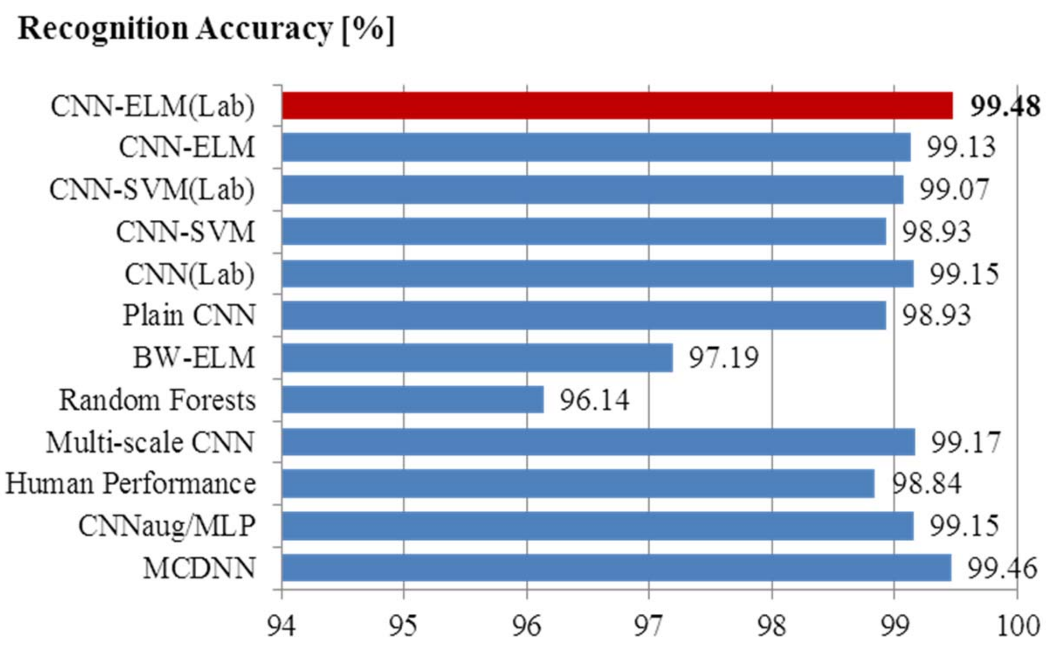
Learning Methods	Testing Accuracy	Training Time
ELM-AE	**86.45**	**602s**
3D ShapeNets (Convolutional Deep Belief Network)	86.5	**Two days**
[K. Xu, et al, 2015]		



Princeton/MIT/CUHK's 3D ShapeNets for 2.5D Object Recognition and Next-Best-View Prediction
[Wu, et al 2014]

Traffic Sign Recognition (DNN + ELM)

Methods	CNN + ELM Based	MCDNN
Accuracies	99.48%	99.46%
Training time	**5 hours (regular PC)**	37 hours (GPU Implementation)
(ELM may just spend several minutes on training in order to reach 98+% accuracy) [Xu, et al 2015]		



Internet of Intelligent Things

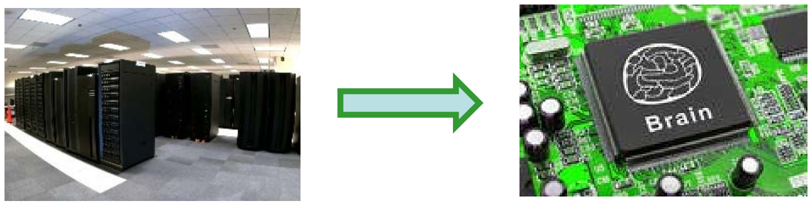
Things

+

ELMs

Intelligent Things

(eg. Intelligent engine, intelligent devices, intelligent sensors, intelligent cameras, etc)



Society of Intelligent Things

Three Stages of Intelligent Things

Internet of Things

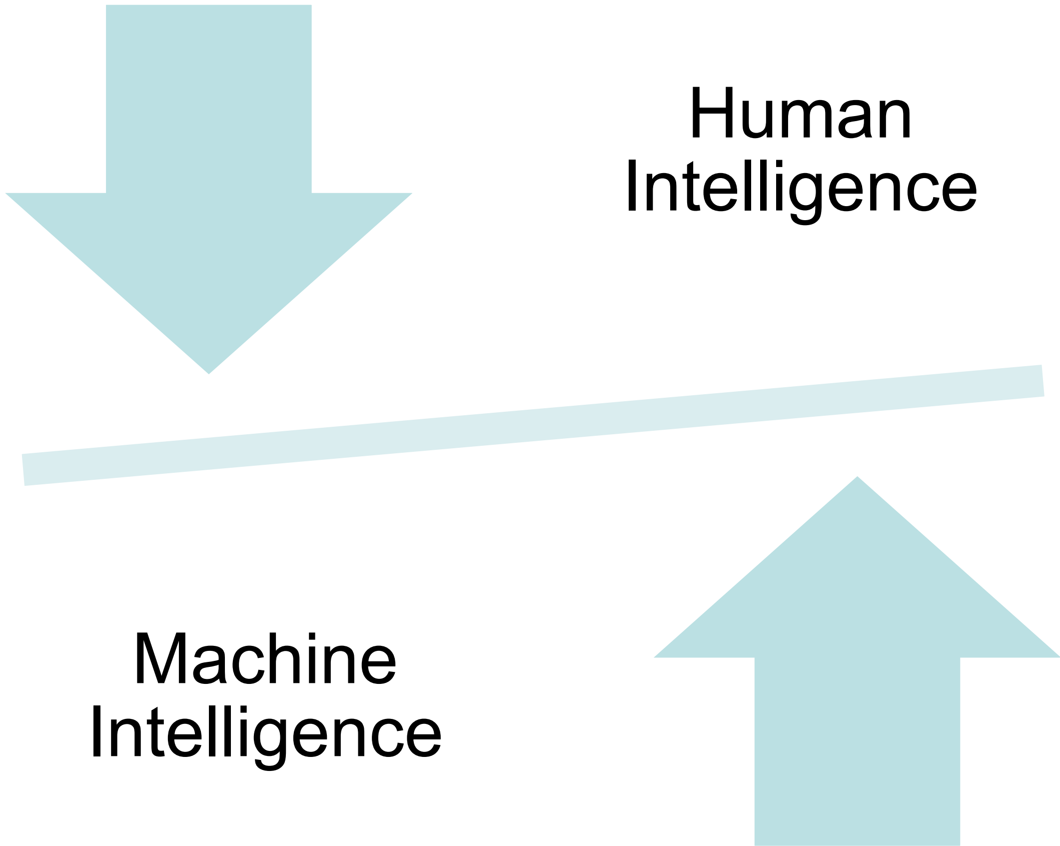
Internet of Intelligent Things

- Intelligent things with ELMs
- Smart materials, smart sensors

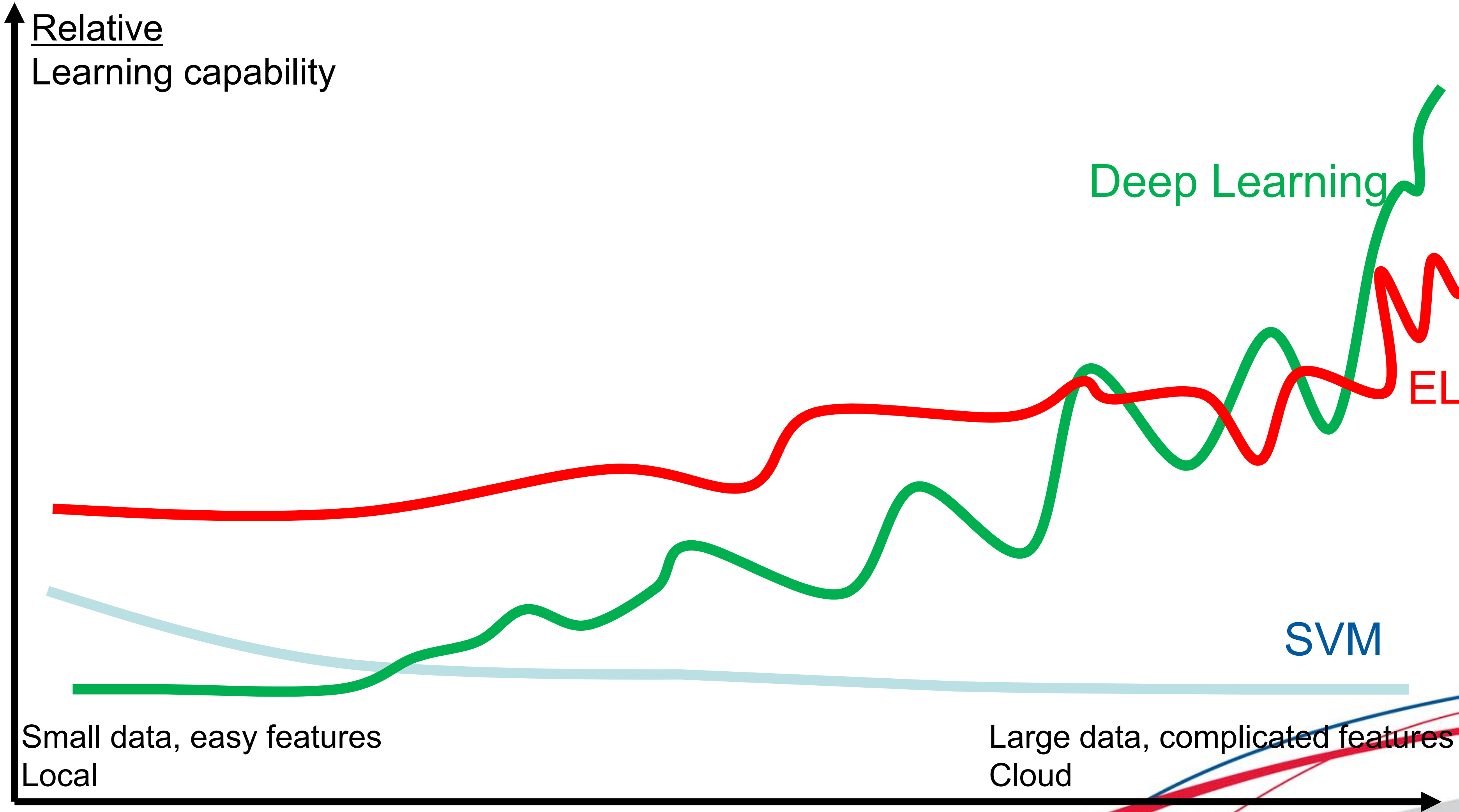
Society of Intelligent Things

- Internet disappearing?
- From living thing intelligence to machine intelligence?

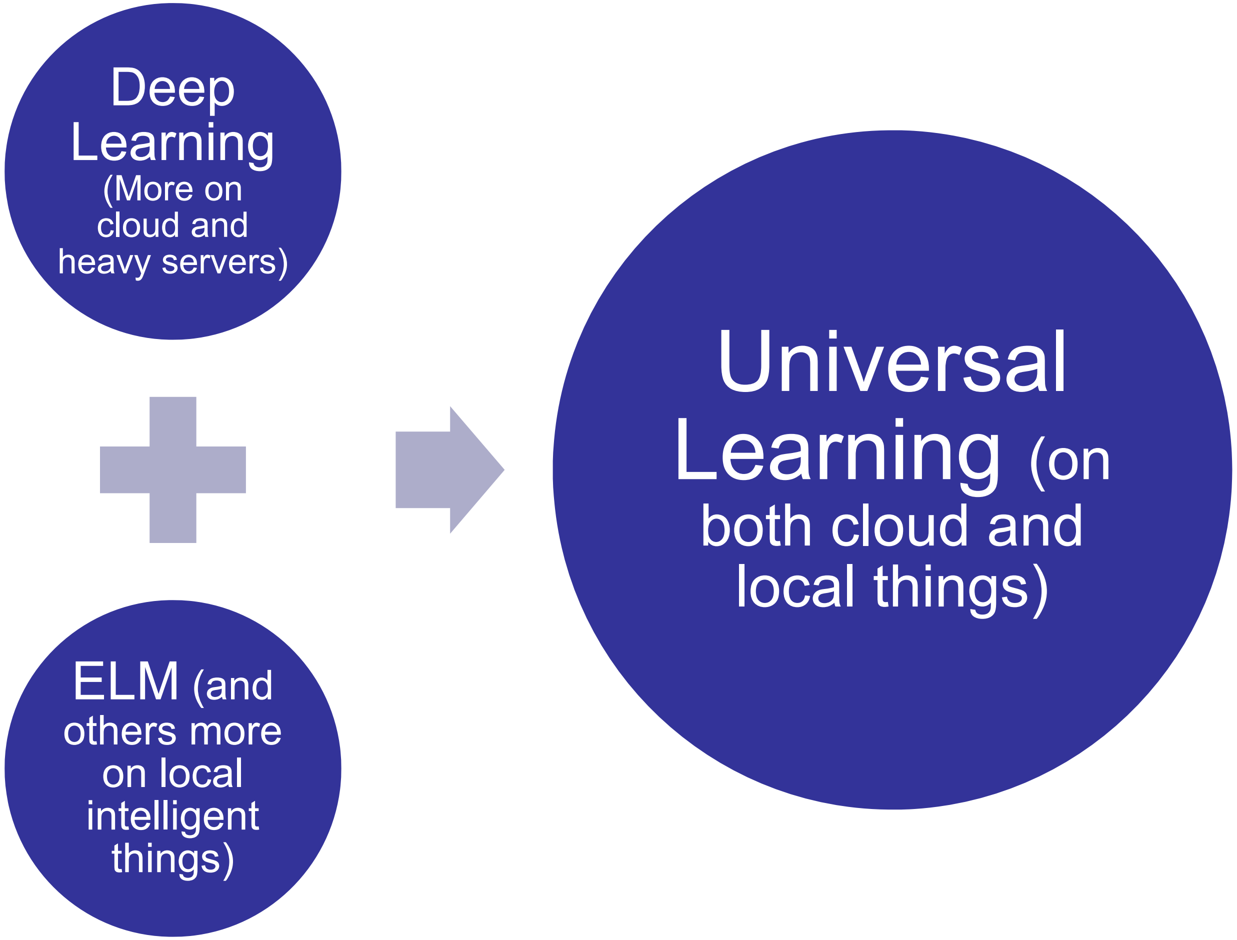
Human Intelligence vs Machine Intelligence



Deep Learning and ELM: Good Combinations



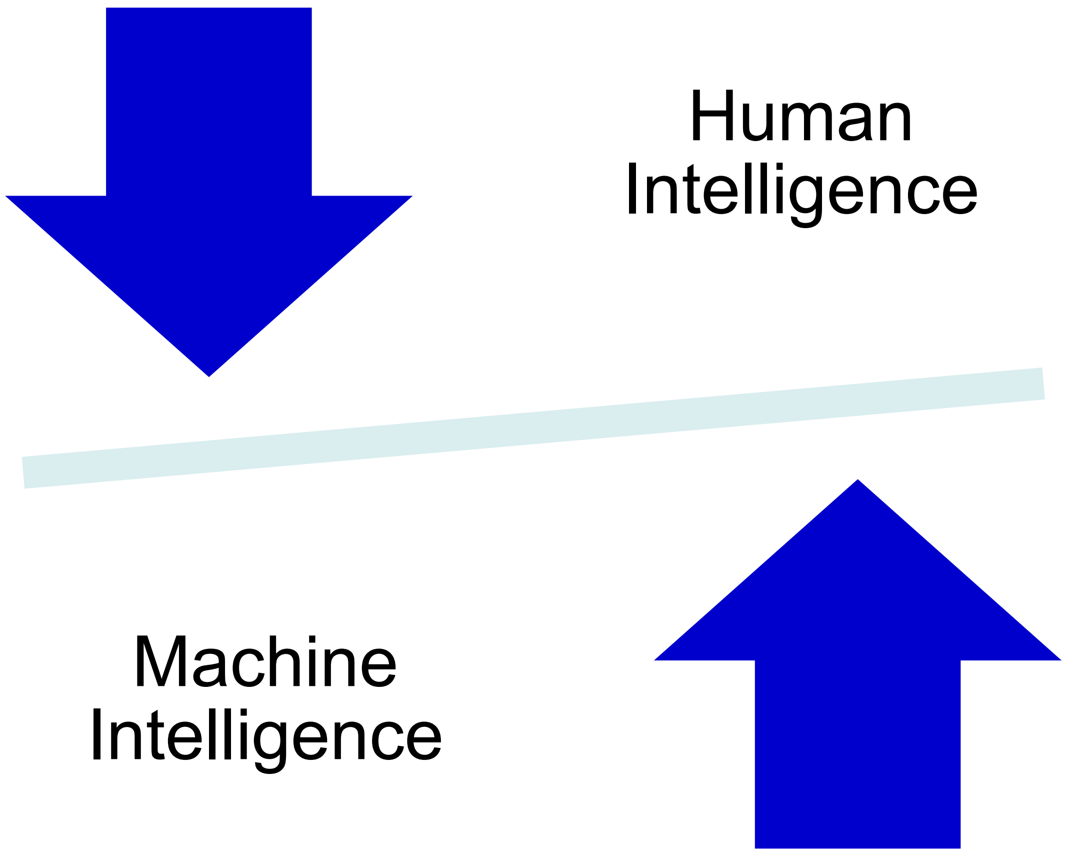
Deep Learning and ELM: Good Combinations



What is Intelligence Index?

Where is FUTURE after Intelligence Revolution?

Machine Intelligence Beats Human Intelligence One By One



Concerns and *Worries*:

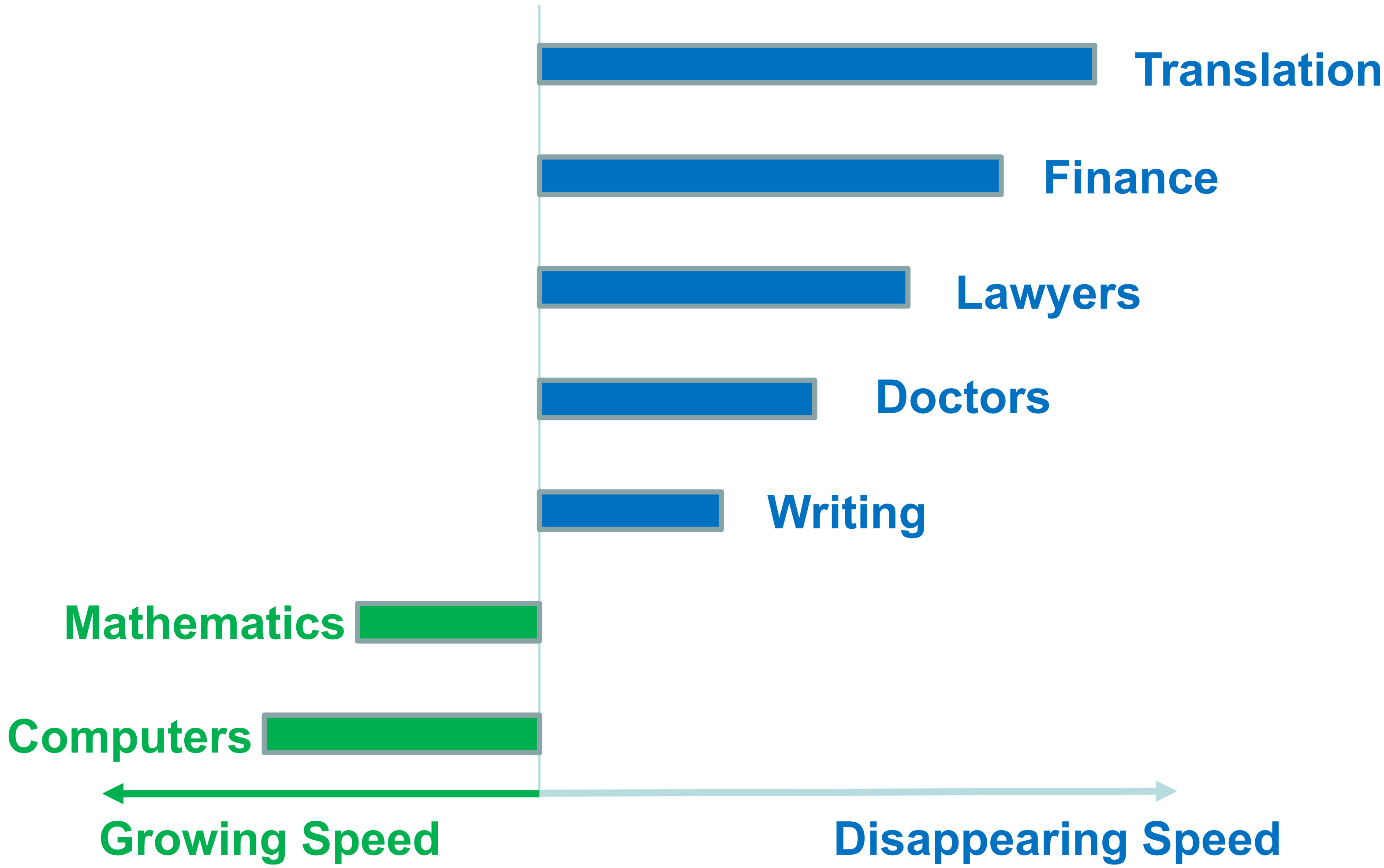
Turning Points for Human Being and Intelligence

Positive and *Negative* Intelligence

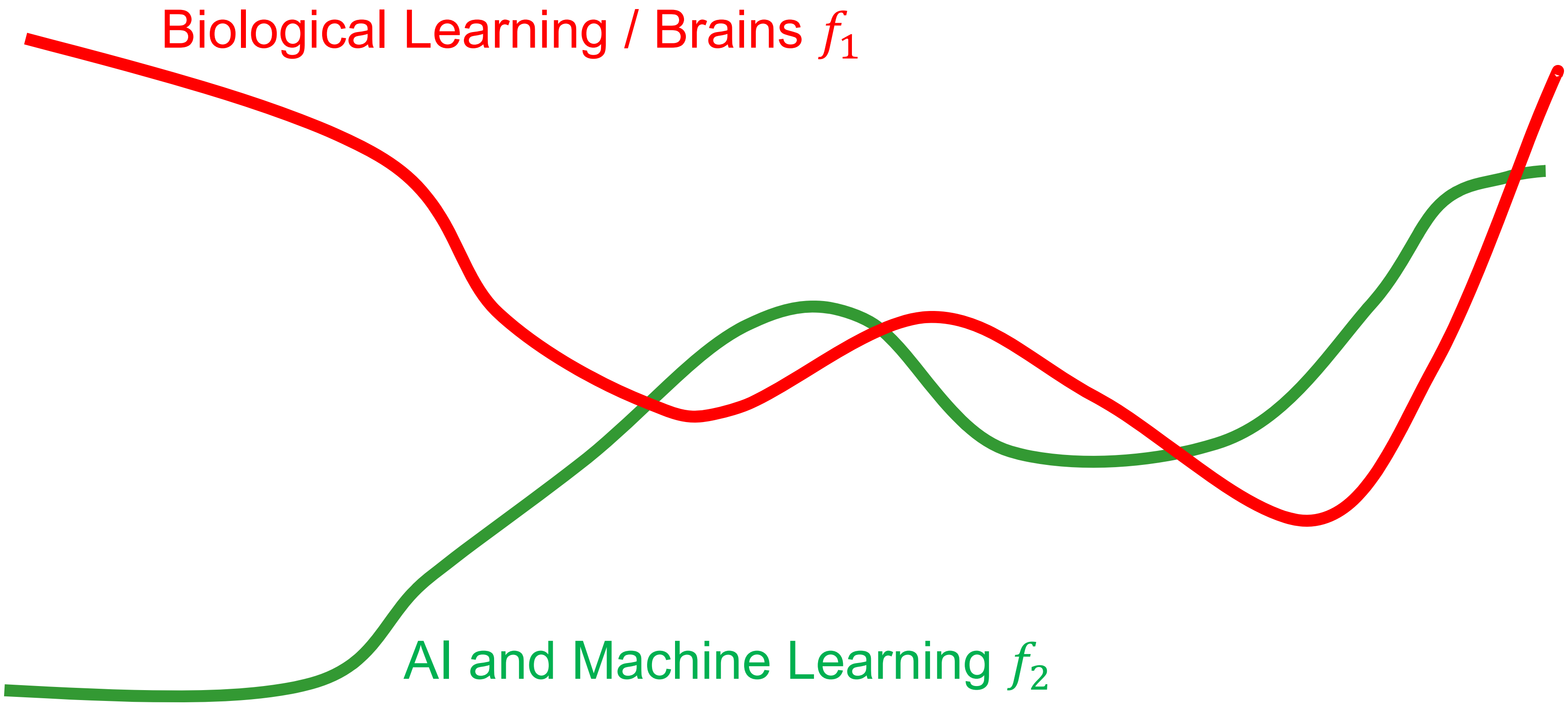
10 Impact of Intelligent Revolution

1. Data driven Engineering and Science (Science, Signal Processing)
2. Divergence of Artificial Intelligence and Machine Learning
3. Intelligence moving from cloud to everywhere, anywhere
4. Smart Material
5. Machine Intelligence and Evolution
6. Intelligent Internet of Things and New Economical Model
7. Bridging Gap of Machine Intelligence and Biological Learning
8. Turning Point of Intelligence
9. Intelligence travelling among universe, reverse of space and timing
10. Ethics and "Control of Intelligence"

Job Impact of Artificial Intelligence



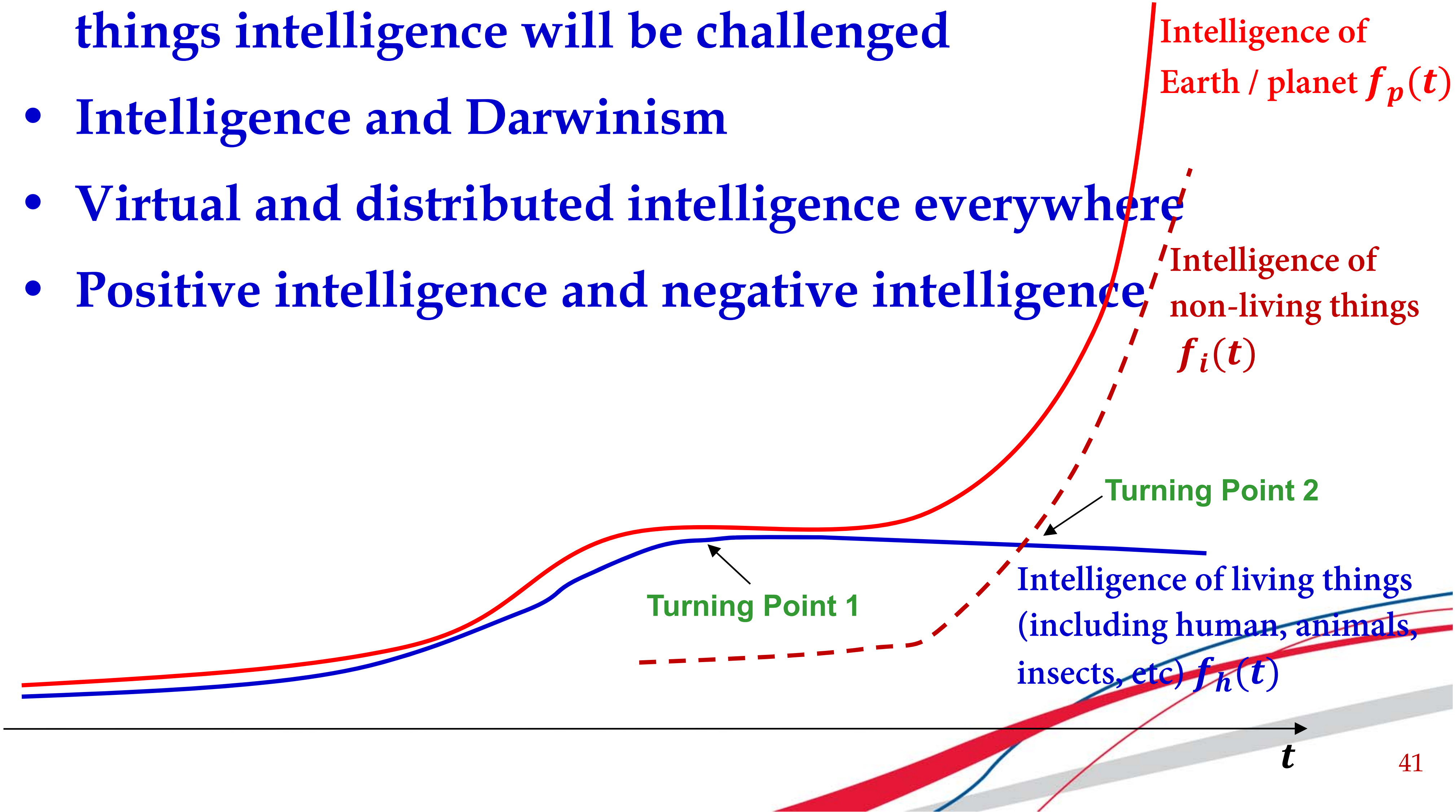
Machine Learning and Biological Learning: Two Learning Functions



Two learning functions converge and diverge in different cases

Intelligence Index

- The entire intelligence of earth is going up sharply.
- The dominant roles of human intelligence and living things intelligence will be challenged
- Intelligence and Darwinism
- Virtual and distributed intelligence everywhere
- Positive intelligence and negative intelligence



Brief History of Living Intelligence