

Historical Trends in Deep Learning

Sargur N. Srihari

srihari@cedar.buffalo.edu

This is part of lecture slides on [Deep Learning](#):
<http://www.cedar.buffalo.edu/~srihari/CSE676>

Logic systems vs Deep learning

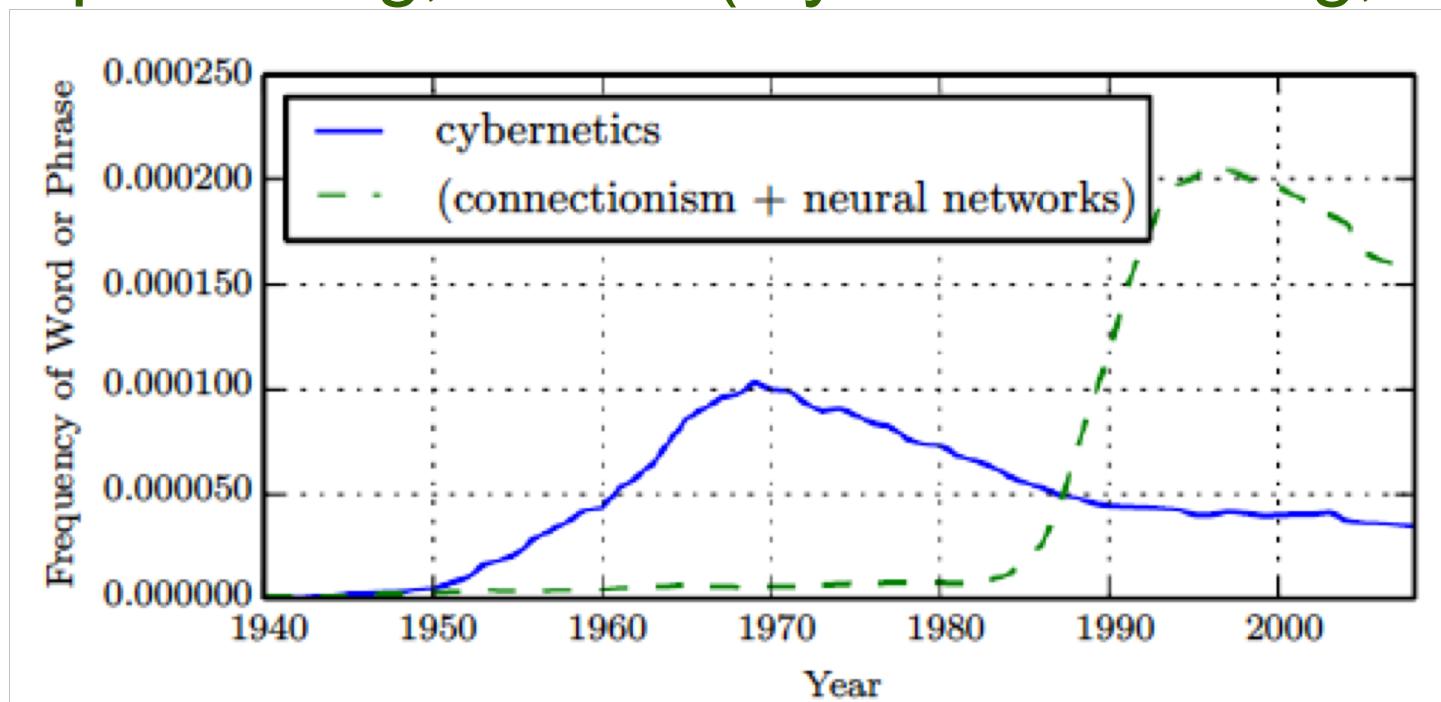
- The difference between failed logic systems and deep learning is *scale*
- The connectionists' neural networks, unlike the rules-based, logic-based approach, were able to scale up to larger and larger problems as computers got more and more powerful and data more plentiful

Time-frame of growth

- People who went for logic had fifty years to show that it didn't scale.
- And now, we had thirty years, from the eighties to today, to show that it [connectionism] does scale.
- "Here, at least with some patterns, with pattern recognition, with reinforcement learning and so forth, we have something that scales,"

History of Deep Learning

- Dates back to 1940s
- Three historical waves:
 - Cybernetics, peaked in 1970
 - connectionism/neural networks, peaked in 1995
 - Deep learning, 2006+ (layerwise training, big data)

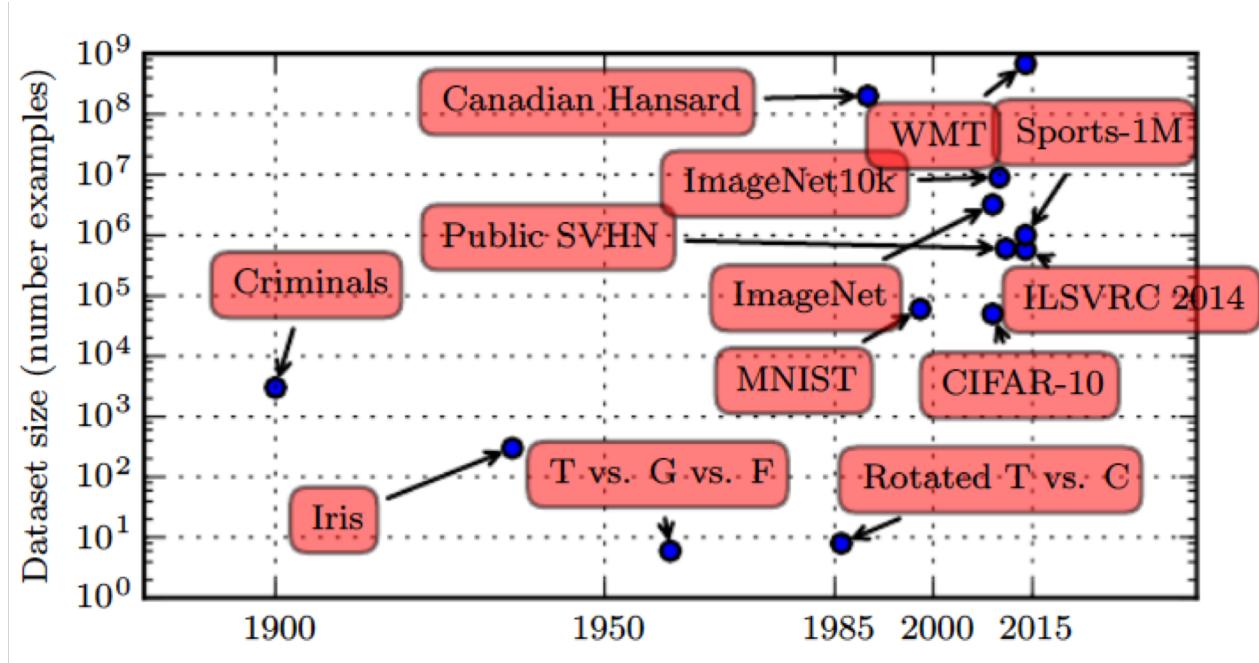


Why this resurgence?

- Data set sizes needed to succeed
- Increasing computational resources (model sizes)
 - Dense connectivity
 - Increased no. of neurons

Large Data Sets

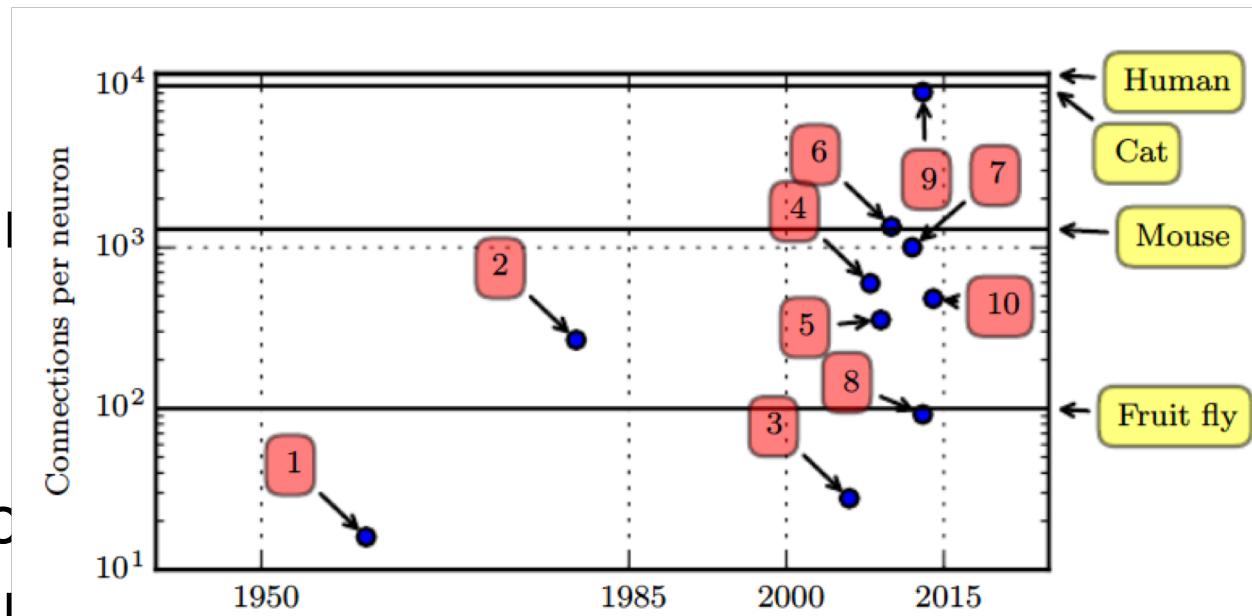
- Algorithms can be provided with data sets needed to succeed



8	9	0	1	2	3	4	7	8	9	0	1	2	3	4	5	6	7	8	6
4	2	6	4	7	5	5	4	7	8	9	2	9	3	9	3	8	2	0	5
0	1	0	4	2	6	5	3	5	3	8	0	0	3	4	1	5	3	0	8
3	0	6	2	7	1	1	8	1	7	1	3	8	9	7	6	7	4	1	6
7	5	1	7	1	9	8	0	6	9	4	9	9	3	7	1	9	2	2	5
3	7	8	2	3	4	5	6	7	8	9	0	1	2	3	4	5	6	7	0
1	2	3	4	5	6	7	8	9	8	1	0	5	5	1	9	0	4	1	9
3	8	4	7	7	8	5	0	6	5	5	3	3	3	9	8	1	4	0	6
1	0	0	6	2	1	1	3	2	8	8	7	8	4	6	0	2	0	3	6
8	7	1	5	9	9	3	2	4	9	4	6	5	3	2	5	5	9	4	1
6	5	0	1	2	3	4	5	6	7	8	9	0	1	2	3	4	5	6	7
8	9	0	1	2	3	4	5	6	7	8	9	6	4	2	6	4	7	5	5
4	7	8	9	2	9	3	9	3	8	2	0	9	8	0	5	6	0	1	0
4	2	6	5	5	4	3	4	1	5	3	0	8	3	0	6	2	7	1	
1	8	1	7	1	3	8	5	4	2	0	9	7	6	7	4	1	6	8	4
7	5	1	2	6	7	1	9	8	0	6	9	4	9	9	6	2	3	7	1
9	2	2	5	3	7	8	0	1	2	3	4	5	6	7	8	0	1	2	3
4	5	6	7	8	0	1	2	3	4	5	6	7	8	9	2	1	2	1	3
9	9	8	5	3	7	0	7	7	5	7	9	9	4	7	0	3	4	1	4
4	7	5	8	1	4	8	4	1	8	6	6	4	6	3	5	7	2	5	9

Increased Connectivity

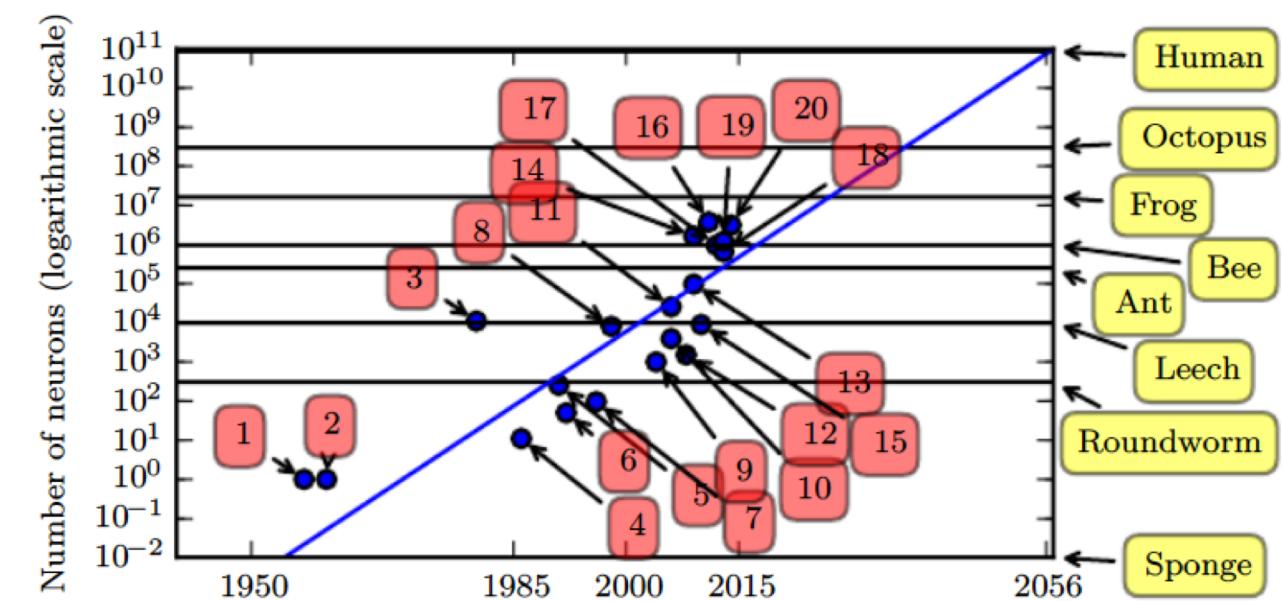
- Intelligence achieved when many neurons work together
1. Adaptive Linear element (Adaline), 1960
 2. Neocognitron
 3. GPU CNN
 4. Deep Boltzmann
 5. Unsup. CNN
 6. GPU MLP
 7. Distributed Auto
 8. Multi-GPU CNN
 9. COTS Unsupervised CNN
 10. GoogLeNet



Increased no. of neurons

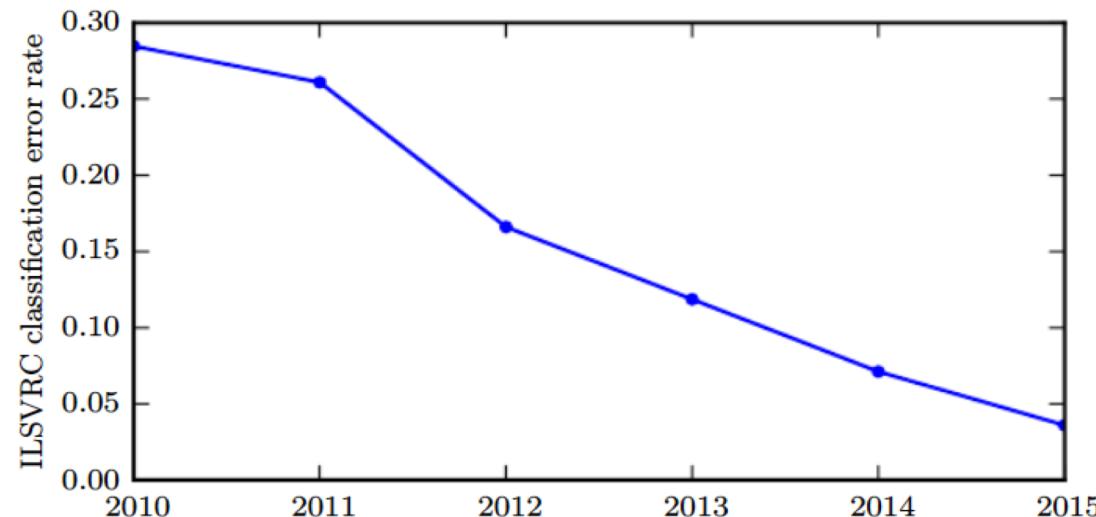
- Since introduction of hidden units, doubling of neurons every 2.4 years

1. Perceptron
2. Adaline
3. Neocognitron
4. Early backprop
5. RNN speech
6. MLP speech
7. Mean-field Belief Net
8. LeNet5
9. Echo-state network
10. Deep Belief Net
11. GPU CNN



Increasing Accuracy

- Image Net Large Scale Visual Recognition Challenge



ILSVRC 2012 tasks

1. Classification
2. Classification with localization
3. Fine-grained classification

