Neural Network

Chapter 1 Introduction to Artificial Neural Network

西安电子科技大学
XIDIAN UNIVERSITY

Outline

01 Artificial Intelligence (AI)

02 Artificial Neural Network (ANN)

03 History of ANN

04 Applications & Prospects of ANN

01

Artificial Intelligence (AI)

Arrangement

- Course No.: Z1HY1025
- Course Class Period: 48 times
- Class Room: 207
- Class Time: Friday 14:00-17:30 pm (3,7-15 week)
  Saturday 14:00-17:30 pm (3,7-15 week)

Course Description

◆Arrangement

- ➢Course Group:
- ➢Course Group No: 593401175

Course Topics

- Chapter One Introduction
- Chapter Two Artificial Neural Network
- Chapter Three Learning & Perceptron
- Chapter Four Optimization
- Chapter Five Convolution Neural Network
- Chapter Six Recurrent Neural Network
- Chapter Seven Reinforcement Learning

Course Description

➢Grading： Homework 30% + Final exam 70%

Final exam	Homework
• Students may not utilize any other materials (in either written or electronic form) during the exams. Students must work independently on the exams and may not utilize electronic communication devices during the exams. Any student violating this policy will receive a grade of zero on the exam.	• Students must work independently on the homework.

◆Intellectual Property

All course materials including but not limited to the syllabus, course assignments, study guides, learning guides, online lecture videos and content, and lab book (i.e. course pack) are property of the instructor and University and may not be shared online or distributed in any manner to others. Students are prohibited from posting course materials or notes online and from selling notes to or being paid for taking notes by any person or commercial firm without the express written permission of the professor teaching this course. Doing so will constitute both an academic integrity violation and a copyright violation. Violations of copyright laws could subject you to civil penalties and criminal liability. Violations of academic integrity may subject you to disciplinary action under University policies.

What is "intelligence"?

- is the general name of wisdom and ability.
- is the basic attribute of human brain.
- connotation: knowledge + thinking
- extension: the ability to discover and apply laws and the ability to analyze and ask questions and solve problems

- **intelligence: ability to understand, learn and reason**
- **intelligent behavior: perception, understanding, reasoning, learning, communication, behavior ability**

What is "Artificial Intelligence"?

- Systems that think like humans. (Haugeland, 1985, Bellman, 1978)
- Systems that think rationally. (Charniak and McDermott, 1985, Winston, 1
- Systems that act like humans. (Kurzweil, 1990, Rich and Knight, 1991)
- Systems that act rationally. (Schalkoff, 1990, Luger and Stubblefield, 1993)

What is "Artificial Intelligence"?

- AI refers to intelligence simulated or realized by computer.
- It is a branch of computer science, which studies the science and technology of how to make machines intelligent. Especially how to realize or reproduce artificial intelligence on a computer.
- It involves many fields such as computer science, brain science, neurophysiology, psychology, linguistics, logic, cognitive science, thinking science, behavioral science and mathematics, as well as information theory, cybernetics and system theory.

Research Goals of Artificial Intelligence

- long-term goal:
  - make intelligent machines with
    - perceptual interaction ability such as reading, listening, speaking and writing
    - thinking ability of association, reasoning, understanding and learning
    - ability to analyze, solve problems and invent
- short-term goal:
  - realize machine intelligence, that is, partial or some degree of intelligence
  - Many achievements have been made: Deep Blue, Backgammon, Watson, Expert System, Auto Drive, AlphaGo, AlphaZero, etc.

Research Methods of Artificial Intelligence

The research on artificial intelligence must draw on the research results of natural intelligence - human brain. According to different emphases, it can be divided into three categories:

- **structural simulation**: neural computing, physiological school, connectionism
- **functional simulation**: symbolic deduction, psychology school, symbolism
- **behavior simulation**: controlled evolution, cybernetic school, behaviorism, evolutionism

Method #1: Structural Simulation - Neural Computing

- physiological school
  - According to the physiological structure and working mechanism of human brai is a local and approximate simulation to realize the intelligence of computer (AI
  - characteristic:
    - use NN's self-learning ability to acquire knowledge, and then use knowledge solve problems
    - It has high parallelism, distribution, strong robustness and fault tolerance.
    - It is good at simulating the image thinking of the human brain, which is convenient to realize the low-level perception function of the human brain: t recognition and processing of images and voices.

Method #2: Functional Simulation - Symbolic Deduction

- psychological school
  - According to the mental model of human brain, knowledge/problems are expressed as some kind of logical network, and the functions of searching, reasoning and learning are realized by symbolic deduction. Such as automatic machine reasoning, theorem proving, expert system, machine game, etc.
  - It is good at simulating the logical thinking of the human brain, so as to realize t advanced cognitive functions (reasoning, decision making, etc.) of the human brain.

Method #3: Behavior Simulation - Control Evolution

- cybernetic school
  - Based on the "perception-behavior model", it simulates the intelligent activities and behavior characteristics of people in the control process: self-optimization, self-adaptation, self-learning and self-organization.
  - It can also be called Situated AI, which emphasizes the interaction between intelligent system and environment.
  - It is believed that intelligence depends on perception and action, and intelligent behavior does not need knowledge. It is believed that human intelligence and machine intelligence can gradually evolve, but they must interact with reality.

Research Fields of Artificial Intelligence

- Computer Vision (CV)
  - Image Recognition, Object Detection, Image Segmentation…
- Natural Language Processing (NLP)
  - Machine Translation, Sentiment Classification, Text Generation…
- Mathematical Theorem Proving
- Data Mining and Knowledge Discovery
- Game Theory
- Robotics
- …

Deep Blue

Deep Blue is a super chess computer produced by IBM in the United States. It weighs 1,270 kilograms, has 32 brains (microprocessors) and can calculate 200 million steps per second. "Deep Blue" has entered more than two million games of outstanding chess players over the past hundred years.

Deep Blue

On May 11th, 1997, "Deep Blue" supercomputer defeated Kasparov, the greatest chess master in human history. This world-famous man-machine war ended with the victory of computers.

The core of Deep Blue algorithm is based on violent exhaustion: all possible moves are generated, and then the search is carried out as deep as possible, and the situation is constantly evaluated to try to find a better way. Comprises a chess moving module, an evaluation module and a search controller.

Watson

In 2011, Watson participated in the variety show Jeopardy to test its ability, and Watson won the first prize of $1 million.

It includes speech recognition, semantic understanding and answering system. First, it need to recognize the voice that you are talking about artificial intelligence. Then it need to understand the questions, because the questions in the corpus may not exactly match the host. Finally, it need to find the corresponding answers in the corpus.

Xiaodu

Xiaodu Robot was born in Baidu Natural Language Processing Department. Relying on Baidu's powerful artificial intelligence, it integrates natural language processing, dialogue system, voice vision and other technologies.

On September 16th, 2014, Xiaodu robot appeared in Jiangsu Satellite TV's open sesame: 40 questions about music, film and television, history and literature were all answered correctly.

In January 2017, I participated in the fourth season of Jiangsu Satellite TV's super brain.

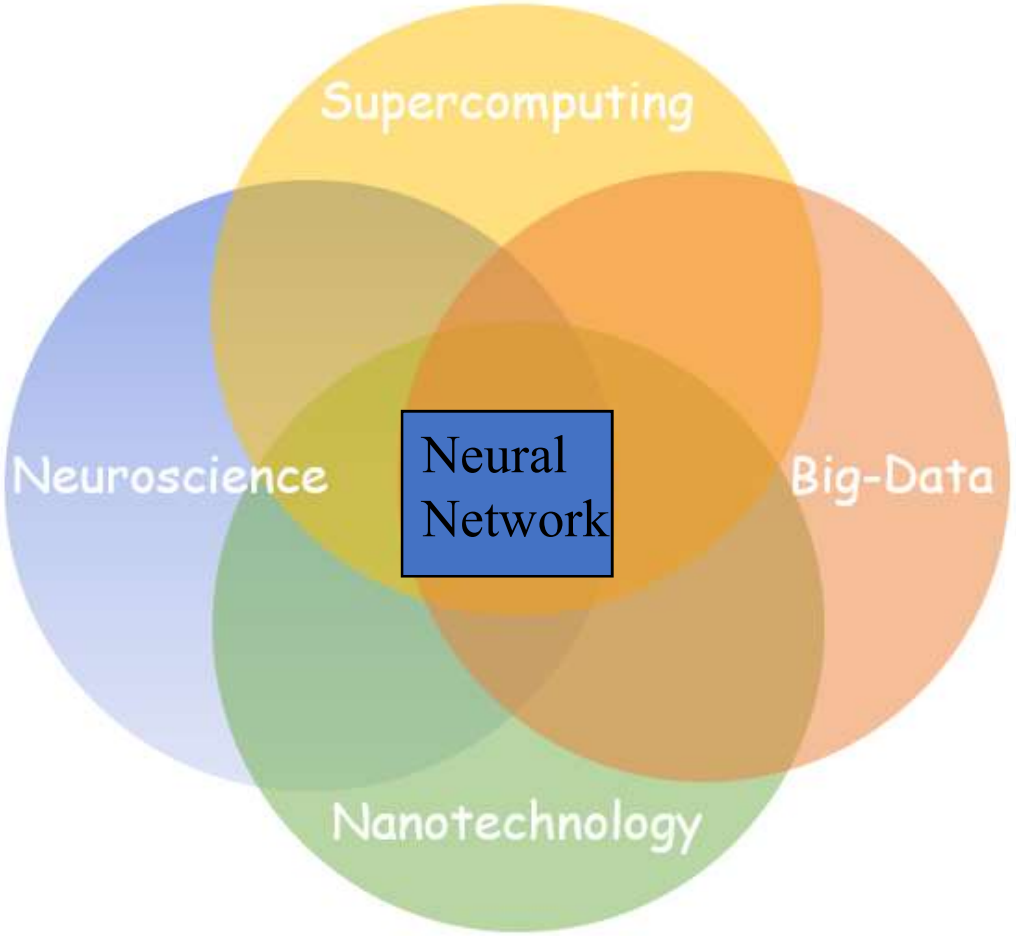
AlphaGo

AlphaGo is developed by DeepMind, a subsidiary of Google. In March 2016, AlphaGo competed with Li Shishi, the world champion of Go and a professional nine-segment player, and won.

At the end of 2016 and the beginning of 2017, the program had a quick chess match with dozens of Go masters from China, Japan and South Korea on the Chinese chess website with "Master" as its registered account. There was no defeat in 60 consecutive games.
In May 2017, he defeated Chinese player Ke Jie.
In October 2017, AlphaGo Zero was released.

core technology: deep learning & intensive learning

Four Combination of Neural Network



Neural Network is combination of principles of **Neuroscience**, **Big Data**, **Supercomputing** and **Nanotechnology**.

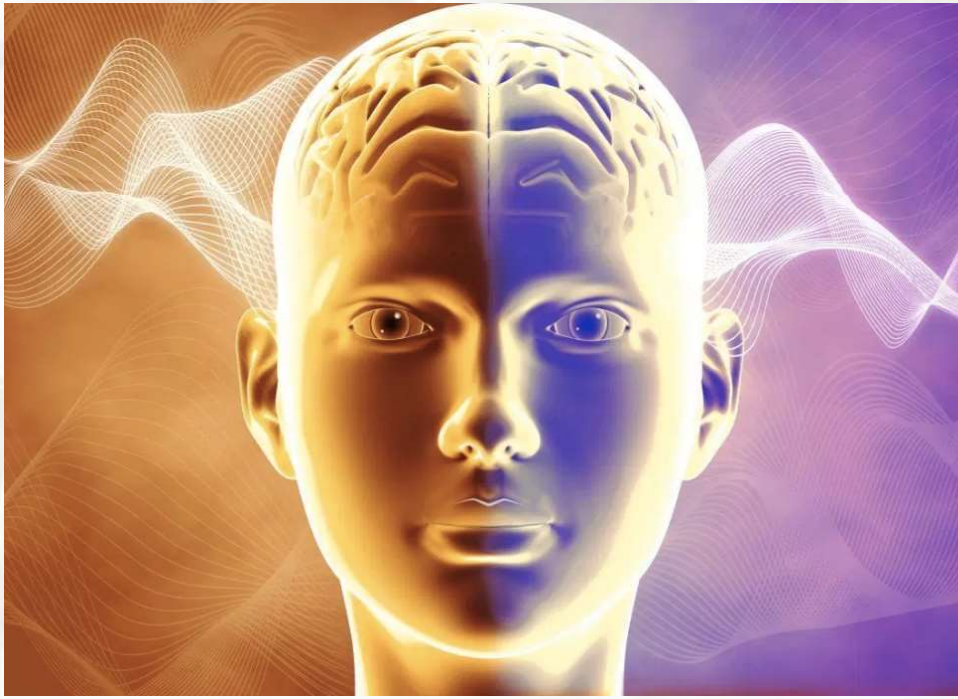
1. Neuroscience

Neuroscience deals with deals with human's thinking process and the function of neural systems.

The architecture of the cognitive computing devices is same as an architecture of brain.

The devices based on this architecture consist of the electronic neurons and synapse and are called Neurosynaptic chips.

There inner network is like brain network.



2. Big data

Big-data provides more than enough information for human and computer to learn, no matter which topic.

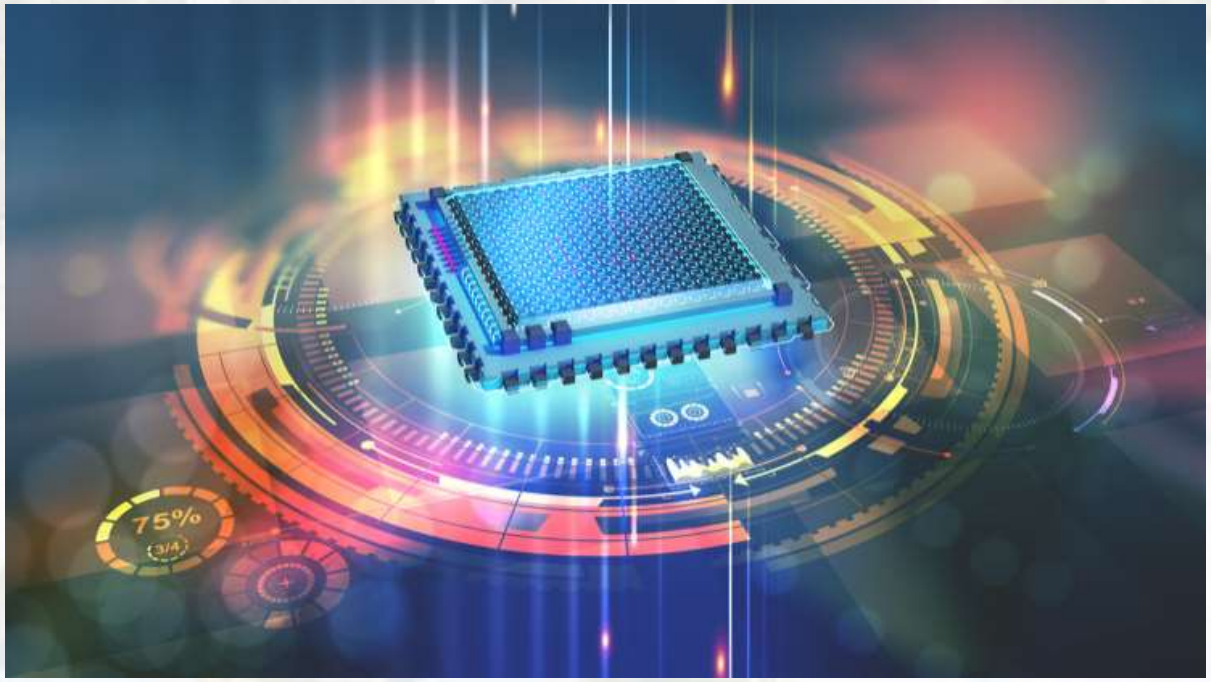
Before the big-data time, very small amount of data can be used for the algorithm to train and test. Thus, big-data enrich the data pool for the learning algorithm.

3. Super Computing

Supercomputing ensures that algorithm and hardware in cognitive computing can achieve high performance.

Without supercomputing, even if we have big data, cognitive computing would be very time-consuming to deal with.

4. Nanotechnology



As we know nanotechnology is scientific field in which material size is 10 to the power -9 meter.

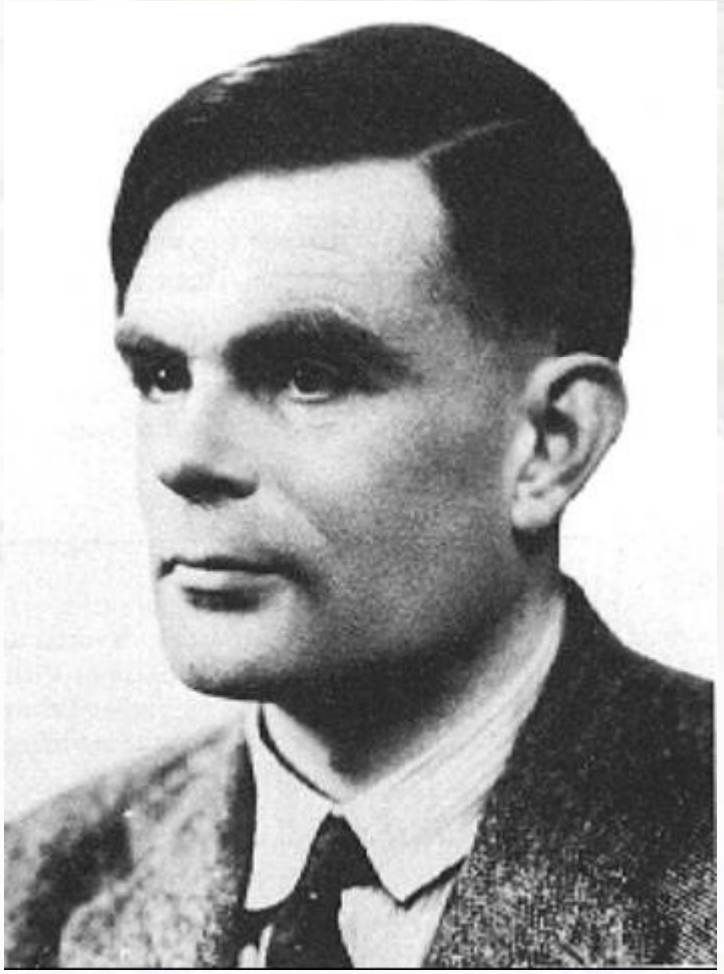
- We need to embed large number of processors and synapse to build the system like the human brain
- So to embed the large number of processor over the chips is done with the help of the nanotechnology.

Masters of Artificial Intelligence in History

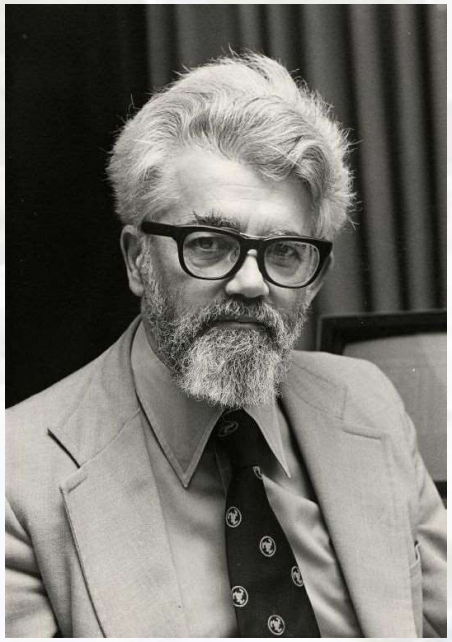
Alan Turing

founder of computer science theory

father of artificial intelligence



Masters of Artificial Intelligence in History

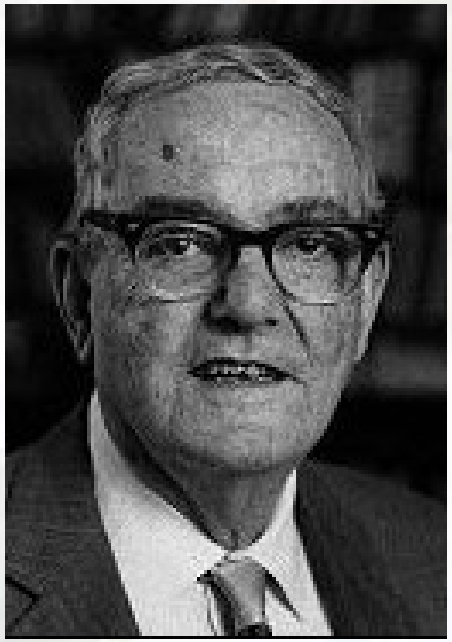


John McCarthy

won the Turing Award in 1971.

Marniv Lee Minsky

won the Turing Award in 1969 (first one to won the award).



Herbert A. Simon

won the Turing Award in 1975.

Allen Newell

won the Turing Award with Simon in 1975.

The Big Four

Geoffrey Hinton
(DNN Research Company)
Google AI Center
University of Toronto

Yann LeCun
Facebook AI Research &
Center for Data Science,
NYU

Andrew Ng
Leader of Baidu Brain &
Google Brain
AI Lab, Stanford

Yoshua Bengio
Canada Research
Chair in Statistical
Learning Algorithms

02

Artificial Neural Network (ANN)

What is "Artificial Neural Network"?

- Artificial Neural Network is abbreviated as ANN.
- Artificial Neural Network is an information processing system which is inspired by biological brain, based on the structure and function of simulated biological brain, and studied by mathematical and physical methods.
- It is a computer system in which many very simple processing units working in parallel are connected to each other in some way and dynamically respond to external input information depending on its state.

What is "Artificial Neural Network"?

- The brain is composed of a large number of nerve cells or neurons. Each neuron can be regarded as a small processing unit. These neurons are connected with each other in some way to form a biological neuron network in the brain. These neurons are excited or inhibited according to the comprehensive size of the received multiple excitation signals.
- The learning process of the brain is a process in which the connection strength between neurons changes adaptively with the external excitation information, and the results of the brain processing information are expressed by the state of the neurons.

What is "Artificial Neural Network"?

- Artificial Neural Network emphasize:
  - ① parallel and distributed processing structure
  - ② The output of a processing unit can be arbitrarily branched with the same size.
  - ③ The output signal can be any mathematical model.
  - ④ complete local operation of processing unit

What is "Artificial Neural Network"?

- Rumellhart, McClelland, Hinton
  ① a group of processing units (PE or AN)
  ② activation state of the processing unit (a_i)
  ③ output function of each processing unit (f_i)
  ④ connection mode between processing units
  ⑤ transfer rule ($\sum w_{ij} o_i$)
  ⑥ the activation rule that combines the input of the processing unit and the current state to generate the activation value (F_i)
  ⑦ modify the learning rules of connection strength through experience
  ⑧ environment of system operation (sample collection)

Characteristics of Artificial Neural Network

- **inherent parallel structure and parallel processing features**
  - The computation function of ANN is distributed on multiple processing units, and the processing units in the same layer operate in parallel. Information processing in ANN is carried out in parallel and hierarchically in a large number of units.
- **distributed storage of knowledge**
  - In ANN, knowledge is not stored in specific memory cells, but distributed among all connection weights throughout the network.

Characteristics of Artificial Neural Network

- **good fault tolerance**
  - When the input is some fuzzy, deformed and other imperfect data and information, ANN can recover the complete memory through association, so to realize the correct recognition of incomplete input information.
- **high nonlinearity and computational inaccuracy**
  - The parallelism of ANN structure and the distributed storage of knowledge i its information storage and processing show the characteristics of spatial distribution and time parallel, which makes the network nonlinear.
  - Because it can deal with some inexact and incomplete fuzzy information, th solution is a satisfactory solution rather than an exact one.

Characteristics of Artificial Neural Network

- **self-learning, self-organization and self-adaptation**
  - Self-learning means that when the external environment changes, after a period of training or sensing, the neural network is able to produce the desired output for a given input.
  - Self-organization means that neural networks can adjust the connection weights by themselves through training, that is, they can adjust the synaptic connections between neurons, so that they have plasticity, and gradually build neural networks that adapt to different information processing requirements.

Alias for Artificial Neural Network

- Artificial Nervous System (ANS)
- Neural Network (NN)
- Adaptive Systems / Adaptive Networks
- Connectionism
- Neurocomputer

Comparison of ANN and Traditional AI Technology

	traditional AI	ANN
basic implementation mode	serial processing; controlled by program	parallel processing; multi-objective learning for sample data; Control is achieved through the interaction between artificial neurons
basic development methodology	design rules, frameworks and procedures; debug with sample data (build a model according to the known environment)	define the structural prototype of the artificial neural network, complete the learning according to the basic learning algorithm through the sample data - automatically extract the connotation from the sample data (automatically adapt to the application environment)
field of adaptation	precise calculation: symbol processing, numerical calculation	imprecise computing: analog processing, sensory, large-scale data parallel processing
simulation object	left brain (logical thinking)	right brain (image thinking)

ANN: Learning Ability

- Artificial neural network can change its behavior according to its environment.
- Different artificial neural network models have different learning / training algorithms.
- self associated network: (auto-associative)
- heterogeneous networks: (hetero-associative)

ANN: Automatic Extraction of Basic Features

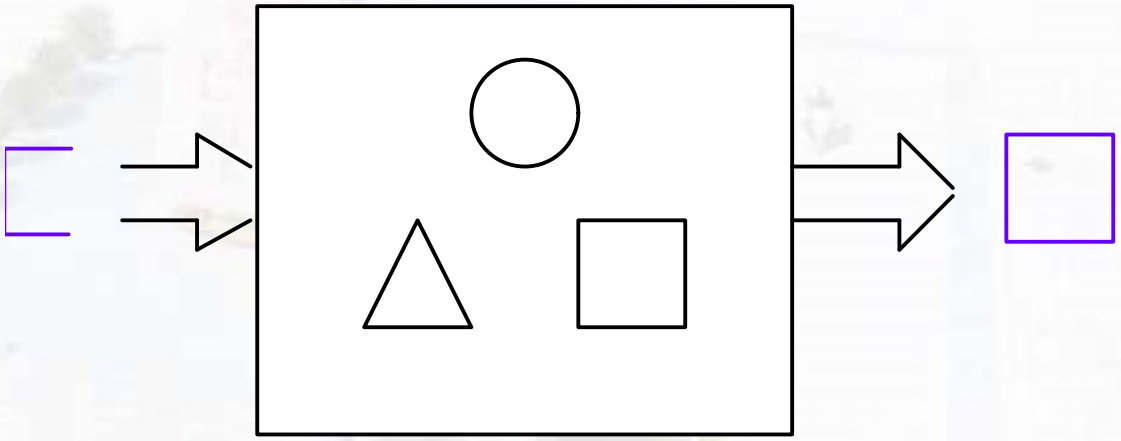
- Because of the imprecision of its operation, it shows the ability of "removing noise and allowing imperfections". By using this imprecision, the automatic classification of patterns can be realized naturally.
- generalization ability and abstraction ability

ANN: Associative Memory

- Because the neural network has the performance of distributed information storage and parallel computing, it has the ability to associate and remember the external stimulus information and input patterns. This ability is realized through the cooperative structure between neurons and the collective behavior of information processing. Neural networks express the memory of information through their synaptic weights and connection structures. This distributed storage enables the neural network to store more complex patterns and recover memory information.

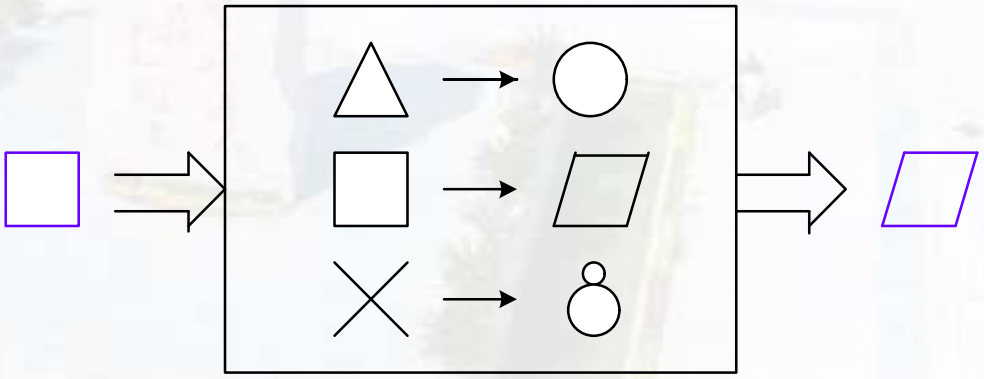
ANN: Associative Memory

- **self associative memory**: The network stores (remembers) a variety of mode information in advance. When part of the information of a stored mode or the information with noise interference is input, the network can recall all the information of the mode through the dynamic association process.

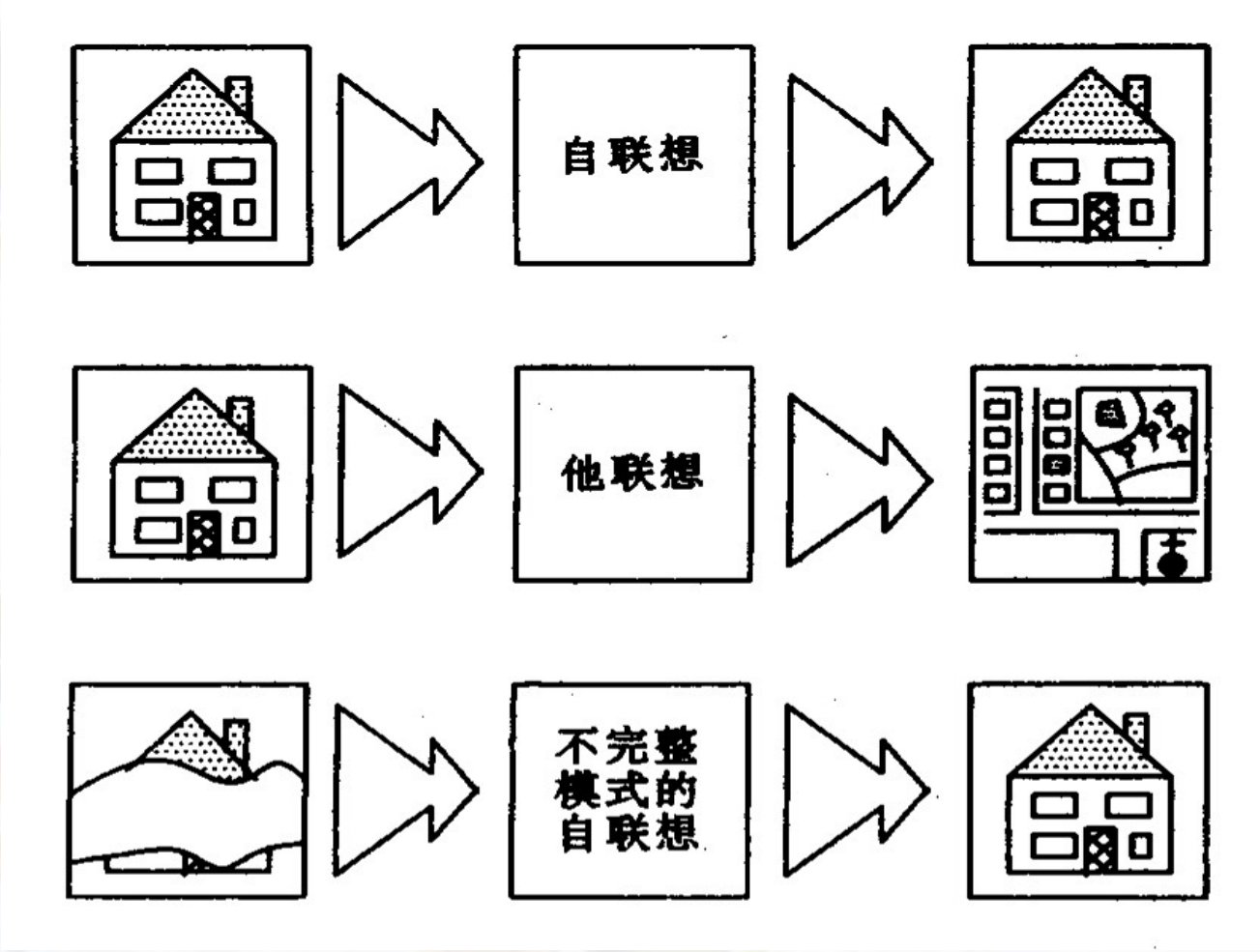


ANN: Associative Memory

- **alloassociative memory**: A plurality of pattern pairs are stored in the network in advance, and each pair of patterns is composed of two parts. When a part of a pattern pair is input, even if the input information is incomplete or noise is added, the network can recall another part corresponding to it.



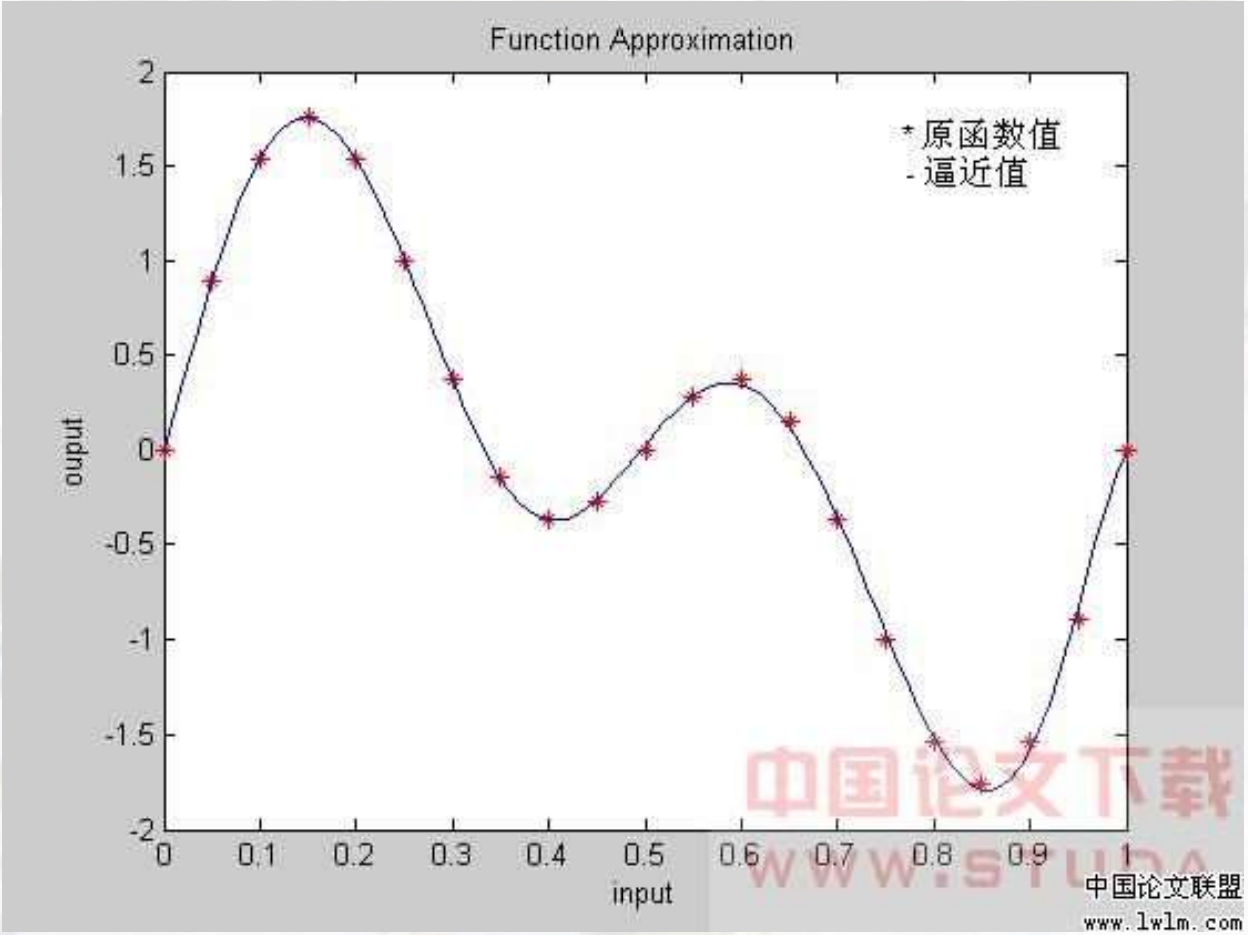
ANN: Associative Memory



ANN: Nonlinear Mapping

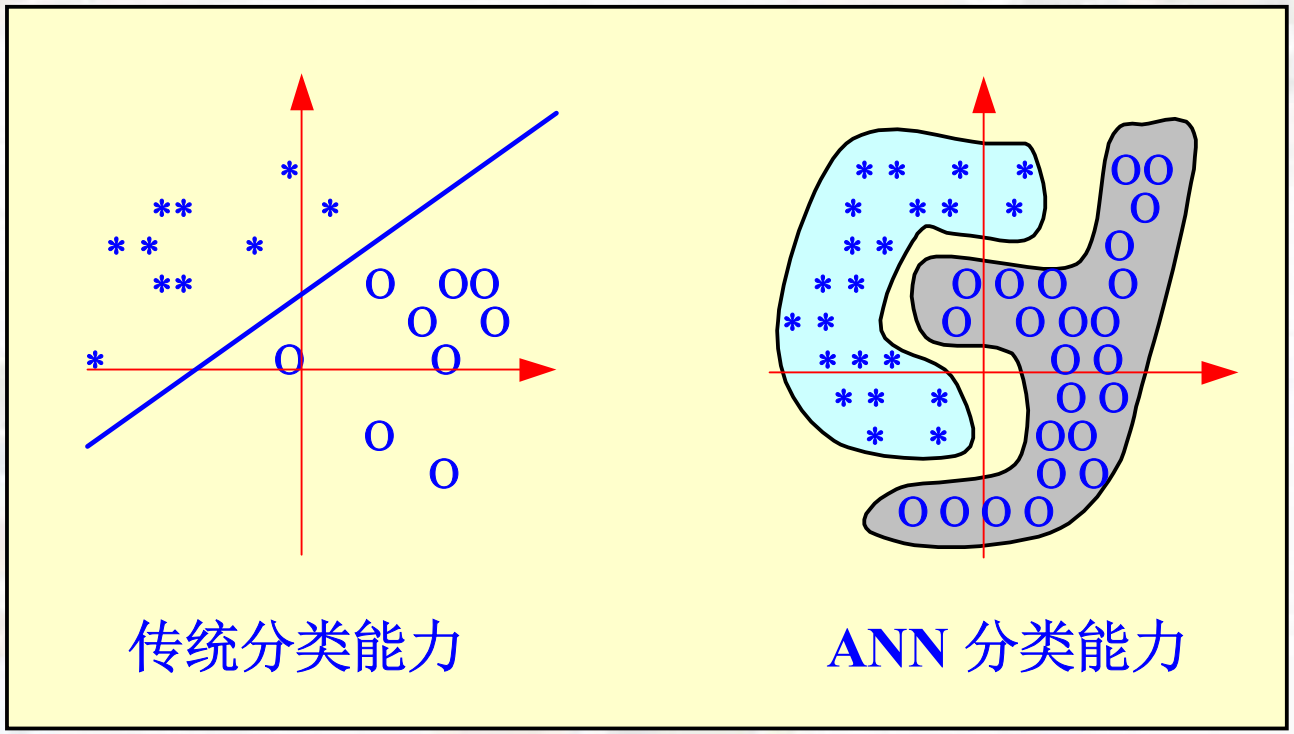
- The designed neural network can approach any complex nonlinear mapping with arbitrary accuracy by automatically learning the input and output sample pairs of the system.

ANN: Nonlinear Mapping



ANN: Classification and Identification

- The classification of the input samples is actually to find the segmentation areas that meet the classification requirements in the sample space, and the samples in each area belong to one class.

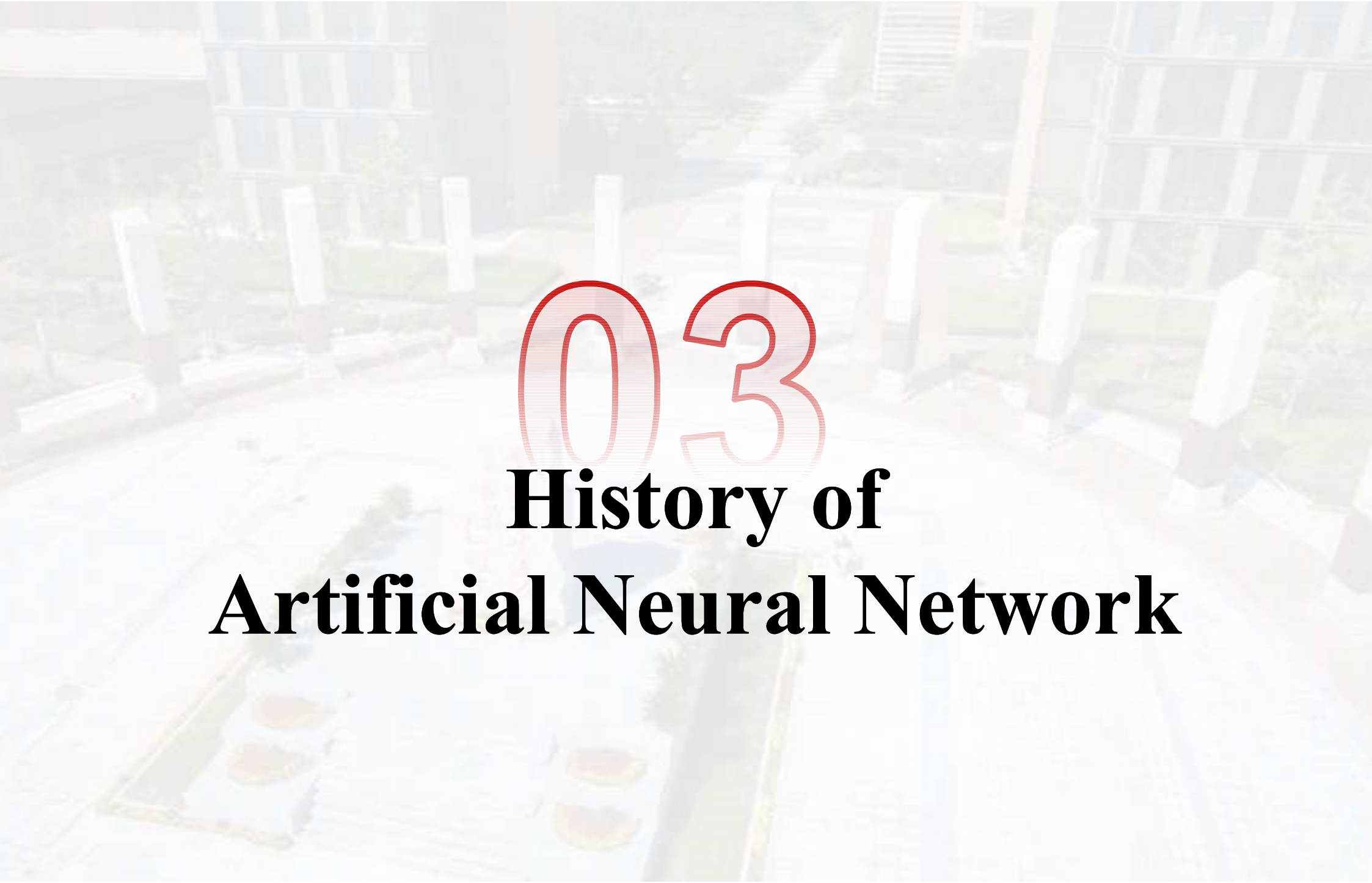


ANN: Distributed Storage of Information

- The distributed storage of information provides fault tolerance.
  - Because the information is distributed and stored in almost the entire network, when a certain point or several points are destroyed, the information can still be accessed.
- The system can still work normally in case of local damage.
- This is not to say that the learning network can be modified arbitrarily. It is also because of the distribution and storage of information that, for a class of networks, if it is allowed to learn new things after it has completed its learning, it will destroy what it has learned.

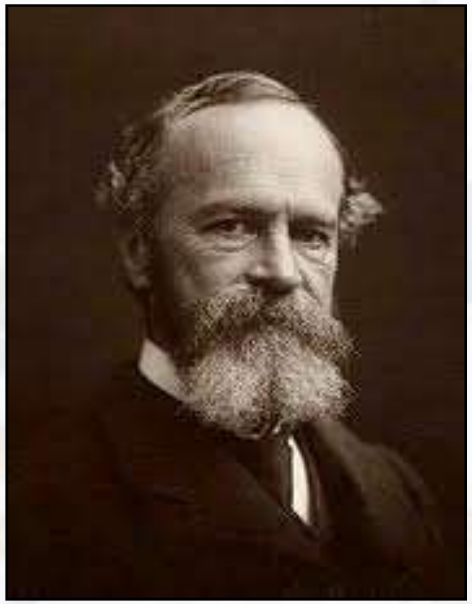
ANN: Adaptability Problem

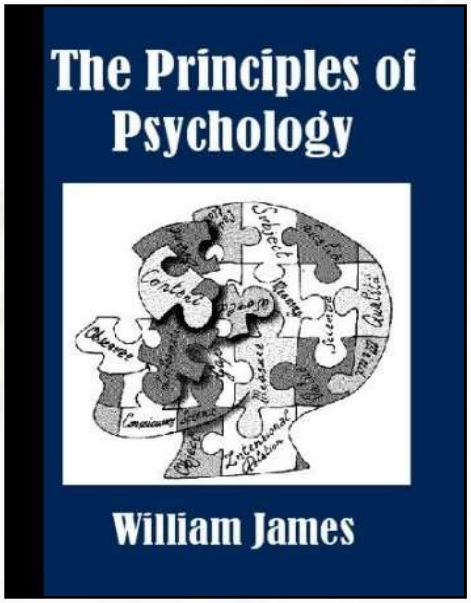
- good at two aspects:
  - A large number of data are classified, and there are only a few cases.
  - It is necessary to learn a complex nonlinear mapping.
- current application:
  - It is mainly used in speech, vision, knowledge processing and decision-making.
  - It also has good applications in data compression, pattern matching, system modeling, fuzzy control, and finding the approximate solution of the optimal solution of combinatorial optimization problems.


03
History of
Artificial Neural Network

Budding Period (1940s)

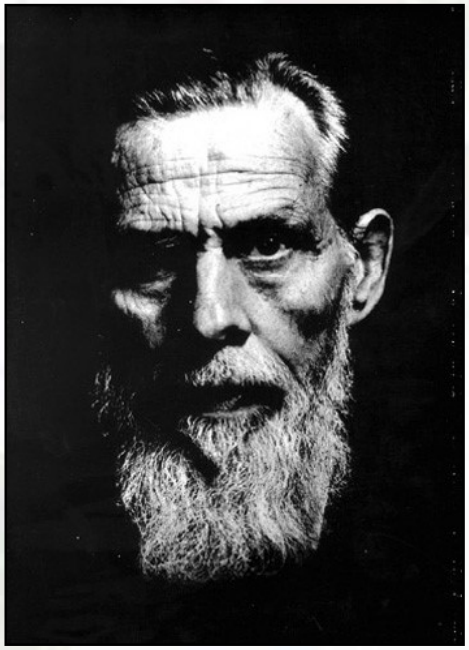
In 1890, American psychologist William James published the first monograph on the structure and function of the human brain, principles of psychology, which made a pioneering study on the basic principles of related learning and associative memory.

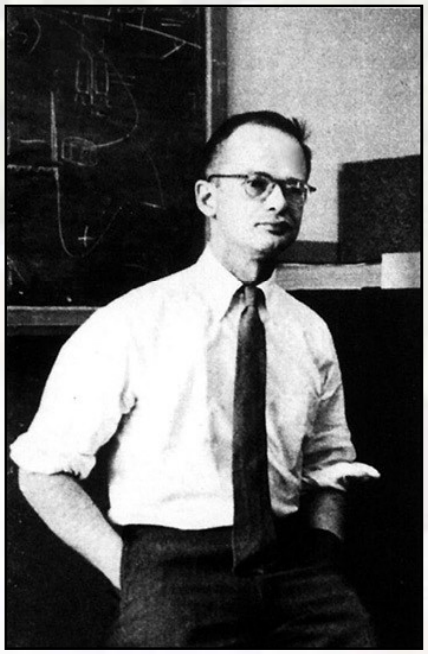




Budding Period (1940s)

In 1943, psychologist McCulloch and mathematical logician Pitts proposed a famous threshold weighted sum model similar to neurons from the perspective of information processing, called M-P model for short. It was published in the bulletin of mathematical biophysics, which ushered in a new era of neuroscience theory.





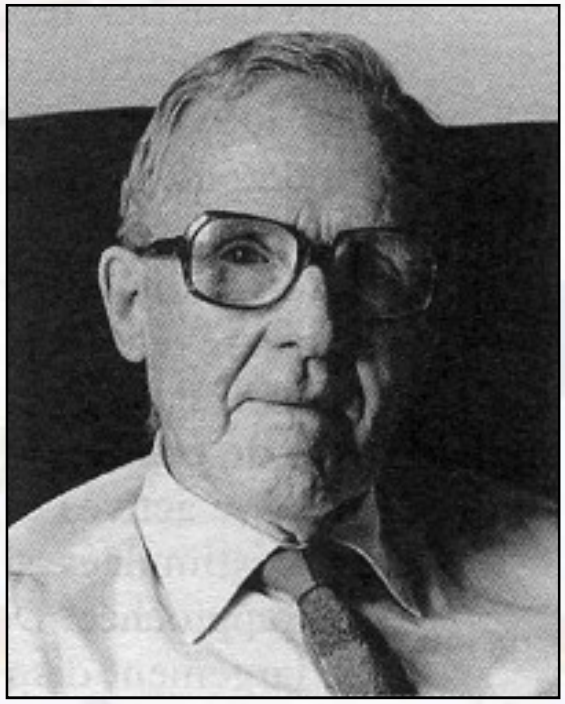
Budding Period (1940s)

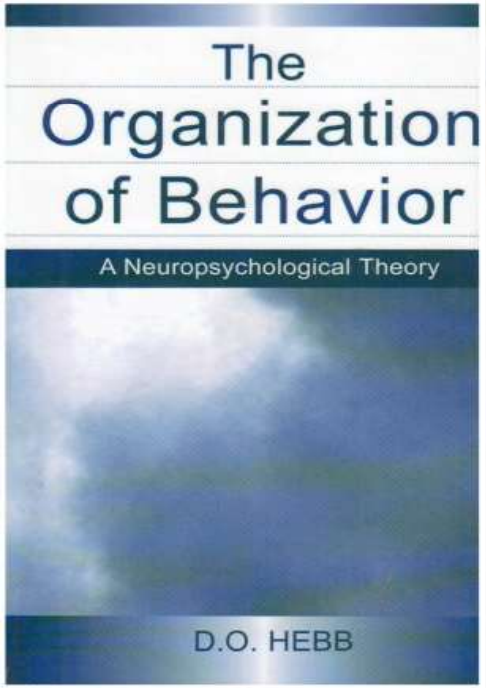
The function of this single neuron model is weak, but the connected network can realize logical operation, including three basic operations: logical multiplication (also known as "and" operation), logical addition (also known as "or" operation) and logical negation (also known as "not" operation).

It pioneered the theoretical research of neural network model and laid a foundation for the research of various neuron models and network models.

Budding Period (1940s)

In 1949, the psychologist Hebb proposed the connection weight training algorithm, namely Hebb algorithm, in the book organization of behavior.





Budding Period (1940s)

Hebb proposed the hypothesis that the strength of synaptic connections between neurons is variable. He believed that the learning process took place at the synapse, and the strength of the synapse changed with the activities of the neurons before and after it. According to this hypothesis, the Hebb rule is proposed to change the strength of neuronal connections. It has great influence on the structure and algorithm of artificial neural network in the future. Hebb's learning algorithm is applied in many artificial neural networks.

First Climax Period (1950~1968)

In 1957, Rosenblatt introduced the concept of Perception, which attempts to simulate the perceptual learning ability of the human brain.

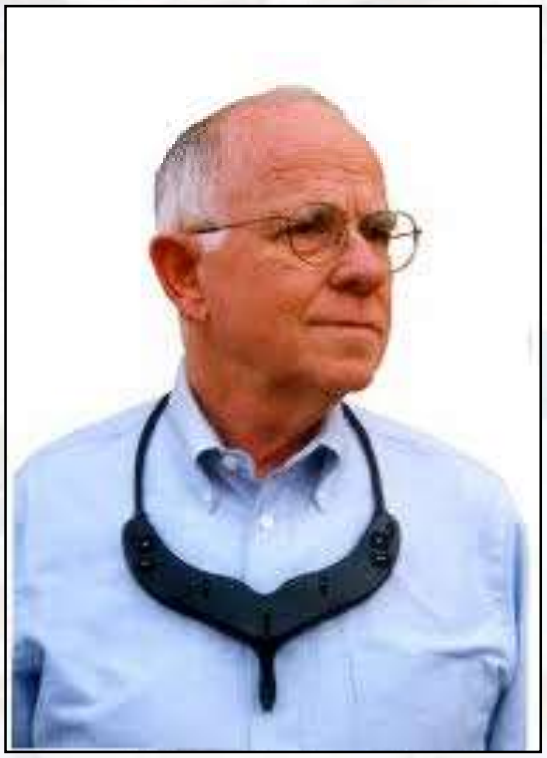
First Climax Period (1950~1968)

The perceptron model proposed by Rosenblatt put the study of neural networks into engineering practice for the first time. It is a psychological model of learning and self-organization, which is essentially in line with the knowledge of neurophysiology. The learning environment of the model is noisy and there are random connections in the network construction, which is in line with the natural environment of animal learning.

This was the first true artificial neural network, and he gave a convergence theorem for two-layer perceptrons. A large class of later neural network models were all deformations of the perceptron model.

First Climax Period (1950~1968)

In 1962, Widrow and Hoff proposed the adaptive linear cell (Adaline), which is a linear network with continuous values.

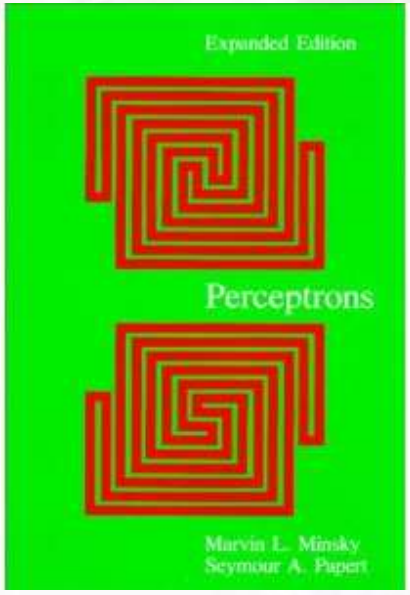


First Climax Period (1950~1968)

- With Marvin Minsky, Frank Rosenblatt, Bernard Widrow and others as representatives, the representative work is the single-level perceptron (Perceptron).
- It can be simulated with electronic circuits.
- There is optimism that the key to intelligence has almost been found. Many sectors have begun to invest heavily in this research, hoping to take the high ground as soon as possible.

First Climax Period (1950~1968)

In 1969, Minsky and Papert, one of the founders of artificial neural networks, published the book Perceptron, which expressed skepticism about the ability of perceptron, so that the research of neural networks was affected, and the research of neural networks went downhill from then on.



First Low Period (1969~1982)

In the late 1960s, Minsky and Papart, the famous American artificial intelligence scholars, conducted an in-depth study of Rosenblatt's work and wrote the influential book "Perceptron", pointing out that the processing power of perceptron is limited, and single-layer perceptron can only make linear division, which will encounter great difficulties for nonlinear or other classification. This is when multilayer neural networks containing hidden units should be used, but finding an effective learning algorithm after introducing hidden units is very difficult, and Minsky asserted that such perceptrons have no scientific research value to speak of, and including multilayer ones makes little sense.

First Low Period (1969~1982)

This conclusion was undoubtedly a heavy blow to the research of neural networks at that time, and objectively played a negative role in the research of neural networks. At the same time, the microelectronics technology at that time also could not provide effective technical guarantee for the research of neural networks. Therefore, in the next decade or so, the number of people engaged in neural network research and financial support dropped greatly, and neural network research was at a low ebb.

First Low Period (1969~1982)

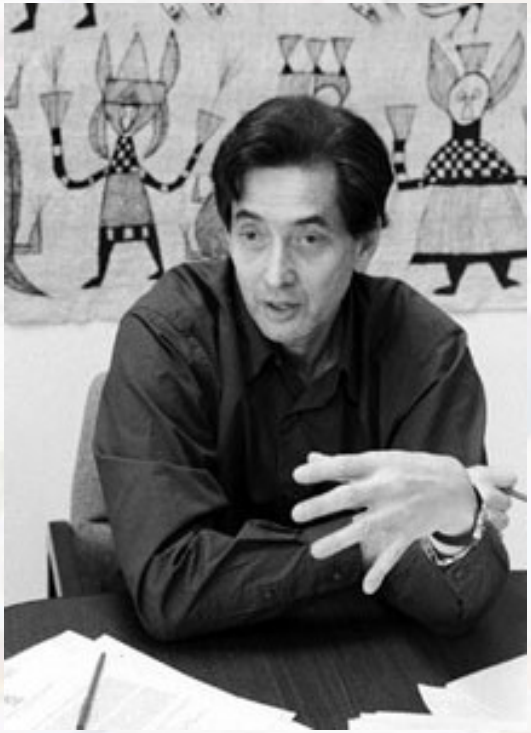
- M. L. Minsky and S. Papert, Perceptron, MIT Press, 1969
- unrepresentable xOR operations (Minsky's challenge)
- research findings from the 1970s and early 1980s
- the law of cognition: knowing - doing - re-knowing

First Low Period (1969~1982)

However, during this period, there were still a few scholars working on neural networks, and Grossberg et al. proposed a theoretical model of adaptive resonance in 1969. 1972 Kohenen proposed a theoretical model of self-organizing mapping and called neural networks as associative repositories. All these theories have laid the theoretical foundation for the further development of neural networks.

First Low Period (1969~1982)

In 1969, American scholars Grossberg and Carperter proposed the Adaptive Resonance Theory (ART) model.



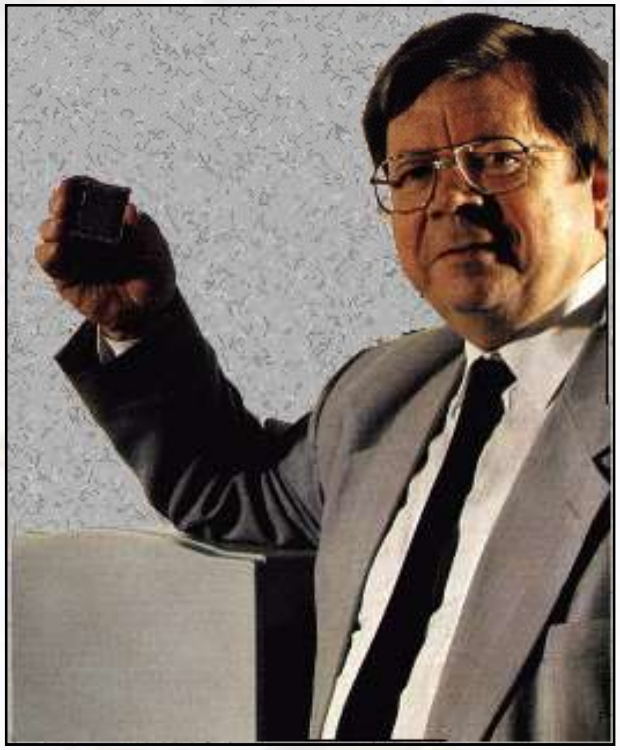
First Low Period (1969~1982)

ART competitive neural network

When we saw a person, if we know the person, we immediately know the person! How should we understand it? In fact, the brain already exists the impression of this person's face, see this person, our brain has a process of search and comparison or recall, if it can search (recall), then congratulations, you recognize this person ... On the contrary, if this person is a stranger, the brain is miserable, searched for half a day, no search (recall), strike, do not recognize. Still a stranger, do not be angry, the brain did not strike, it has silently stored this person's face! If you subsequently have more contact with this person, such as meeting again or talking or interacting together, the brain's memory of this person will gradually strengthen! In actual life, we may have the feeling that we have met someone somewhere, but this feeling is actually generated by the fact that the brain's memory of this person is not strong enough!

First Low Period (1969~1982)

In 1972, the Finnish scholar Kohonen proposed the theory of self-organizing mapping (SOM).



First Low Period (1969~1982)

The research in brain science has shown that there is extensive self-organization of cell populations in the human cerebral cortex. Neurons in different regions have different functions, they have different characteristic input information patterns, and they are sensitive to input signals from different sensory input patterns, thus forming different perceptual pathways in the brain. These neuronal properties are not exclusively biologically inherited, but depend heavily on acquired learning and training.

First Low Period (1969~1982)

Self-Organizing Feature Maps

Self-Organizing Feature Maps (SOM) is based on this theory and has become the most widely used method for self-organizing neural networks. When a neural network receives external input patterns, it will be divided into different response regions, and each region has different response characteristics to the input patterns. This network simulates the function of self-organized feature mapping of the brain nervous system. It is a competitive neural network, and unsupervised learning algorithm is used for network training. This network is widely used in sample classification, sorting and sample detection.

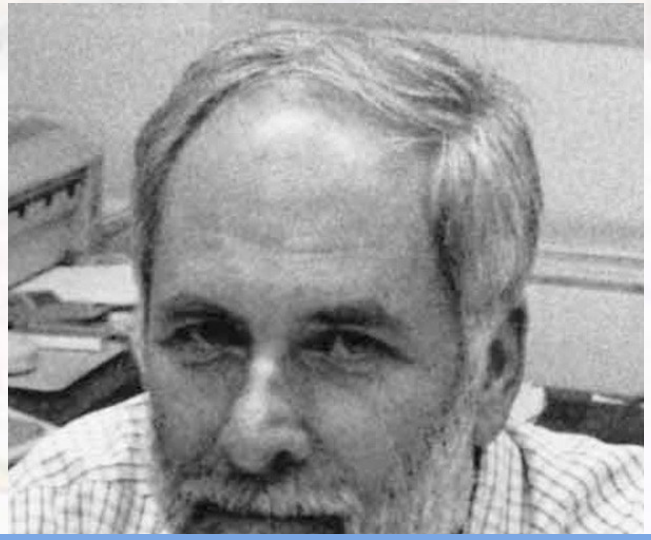
First Low Period (1969~1982)

In 1979, Fukushima proposed the Necognitron (cognitive machine).

The cognitive machine (Neocognitron), proposed by Fukushima in 1972, is by far the most complex structured multilayer network with selective attention capability through tutorless learning and insensitive to smoothness and rotation of the sample. The disadvantage is that the parameters are not easy to choose. It is mainly used for character recognition.

First Low Period (1969~1982)

In 1977, neuropsychologist Anderson (Anderson) proposed the BSB (Brain-State-in-a-Box) model.



The BSB model is a single-layer network with lateral connections between nodes and nodal self-feedback, available with the most self-associative off-neighborhood classifier and can store any analog vector pattern.

First Low Period (1969~1982)

In 1974, Werbos proposed the BP theory, which laid the foundation for the development of neural networks.

Second Climax Period (1983~1990)

In 1982, Hopfield, a physicist at the California Institute of Engineering, published a paper in the Proceedings of the National Academy of Sciences proposing a new approach to associative memory and optimal computation -- the Hopfield model.

Second Climax Period (1983~1990)

In 1982, J. Hopfield proposed recurrent networks

- establishing the discriminatory basis of ANN stability
- clarified the relationship between ANN and dynamics
- a nonlinear dynamics approach to study ANN properties
- point out that the information is stored on the connections of neurons in the network

$$V = \frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n} w_{ij}s(x_i)s(x_j) - \sum_{i=1}^{n}\int_0^{x_i} s_i'(\theta_i)\beta_i(\theta_i)d\theta_i - \sum_{i=1}^{n}\sum_{j-1}^{n} w_{ij}$$

Second Climax Period (1983~1990)

In 1984, Hopfield's model was modified and a hardware principle model for artificial neural networks was proposed using the basic components of analog circuits, which laid the foundation for the implementation of hardware. 1985 Hopfield and Tank proposed the use of neural networks to solve the TSP combinatorial optimization problem.

Second Climax Period (1983~1990)

The action principle of the Hopfield model is that as long as the state of the neural network, determined by the algorithm of neuronal excitation and the strength of binding between neurons, has not yet reached stability for a given excitation pattern, the state will continue to change until a predefined energy function that must decrease reaches a minimal value, when the state reaches stability and no longer changes.

Second Climax Period (1983~1990)

In 1984, Hopfield designed and developed a circuit for his proposed neural network model and showed that each neuron in the network could be implemented with an operational amplifier. He simultaneously conducted research on the application of neural networks and successfully solved the traveling salesman problem (TSP) , which caused a shock to the world. These results brought the research on neural networks back to a new flourishing period.

Second Climax Period (1983~1990)

In 1985, researchers in the parallel distribution processing (PDP) group of Sejnowsky, Rumelhart, Hinton, and others at UCSD introduced a stochastic mechanism in Hopfield networks, proposing the so-called Boltzmann machine.

Second Climax Period (1983~1990)

In 1986, Rumelhart and McClelland proposed the error back propagation learning algorithm (BP algorithm) for multilayer network learning, which better solved the learning problem of multilayer networks.

Second Climax Period (1983~1990)

- 1988, RBF design neural network.
- In the early 90s, Vapnik proposed SVM.
- The first domestic conference on neural networks was held in Beijing in December 1990.

Second Low Period (mid 1990s)

In the mid-1990s, with the rise of statistical learning theory and support vector machines, the weaknesses of the theoretical nature of neural network learning in terms of lack of clarity, trial and error, and the abundance of tricks in its use became more apparent, so that neural network research went into another slump, and the NIPS conference did not even accept papers on the topic of neural networks for many years.

Re-Awareness and Applied Research Period (1991~)

- There are problems such as the application surface is not wide enough and the results are not accurate enough.
- improve existing model models, algorithms to improve the training speed and the accuracy of the network operation
- integration of algorithms
- looking for new breakthroughs in theory and building new dedicated/generic models and algorithms
- Further research on the biological nervous system continues to enrich the understanding of the human brain.

Third Climax Period

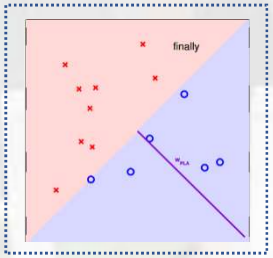
- In 2006, Hinton published an article on deep neural networks in Science.
- In 2010, the U.S. Department of Defense DARPA program funded DNN research at Stanford University, New York University and the NEC American Institute.
- In 2012, the Google Brain project used 16,000 CPUs to build a DNN platform, which made important progress in the field of speech and image recognition.
- Fully automated simultaneous interpretation using DNN technology at Microsoft Research Asia, 2012.
- In 2013, MIT Technology Review listed DNNs as one of the top 10 breakthrough technologies of 2013.

Third Climax Period

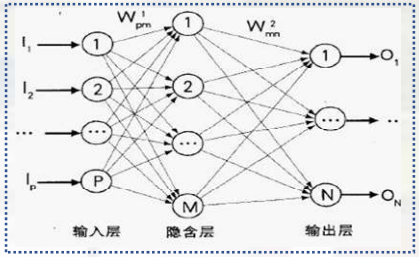
- Facebook's Deepface project in 2014 resulted in a recognition rate of 97.25% for face recognition technology.
- In 2015, the China Brain Project was approved by the State Council and listed as a "major science and technology project for the future development of China".
- 2015, Tsinghua, Institute of Automation, Fudan, "brain-like computing", "brain-like intelligence".
- The victory of DNN-based AlphaGo over Lee Seok in 2016.

History of Artificial Neural Network

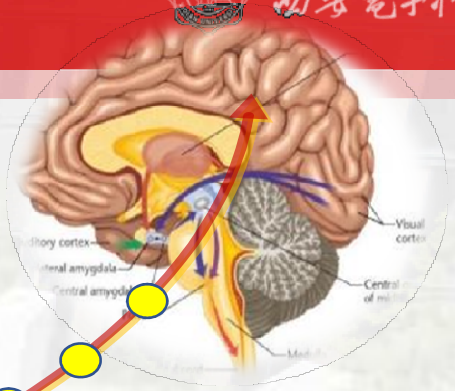
Donald Hebb
Psychology / Mathematician
Hebb Learning Rules
1949



F.Roseblatt
Perceptron
1957
(first climax period)



Rumelhart
BP algorithm
1986

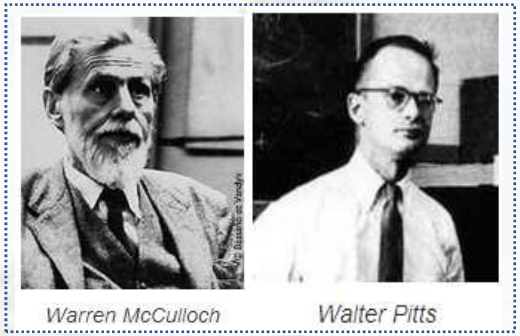


DBN, CNN, AE
VAE, RNN, GAN
2006

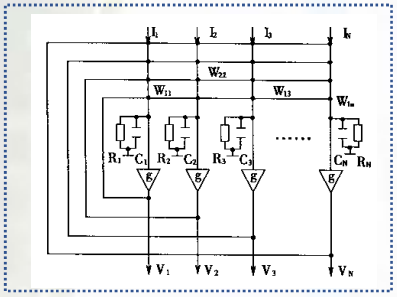
W.S.MeCulloch/W.Pitts
MP model
1943

(Minskey & Papert)
1970s, low period

Hopfield
Hopfield Model
1982
(second climax period)



Shallow model



Deep model

third climax period?

04

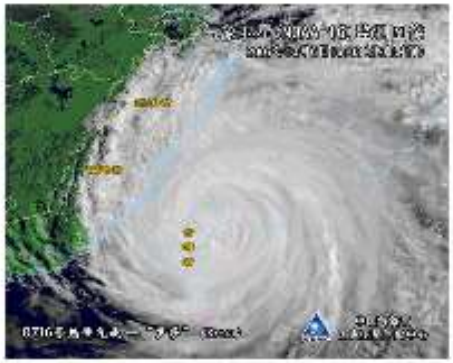
Applications & Prospects of Artificial Neural Network

Appliation Areas of Artificial Neural Network

图片

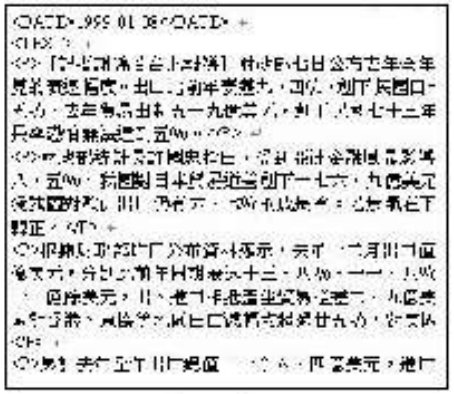
电视

视频监控



遥感图像

语音



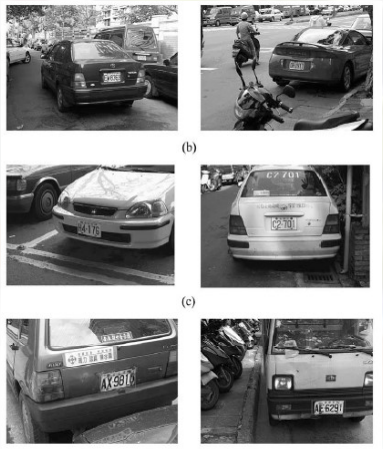
文本

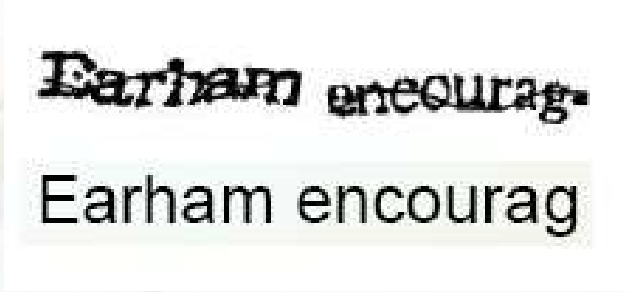
网络数据

医学图像

Appliation Areas of Artificial Neural Network

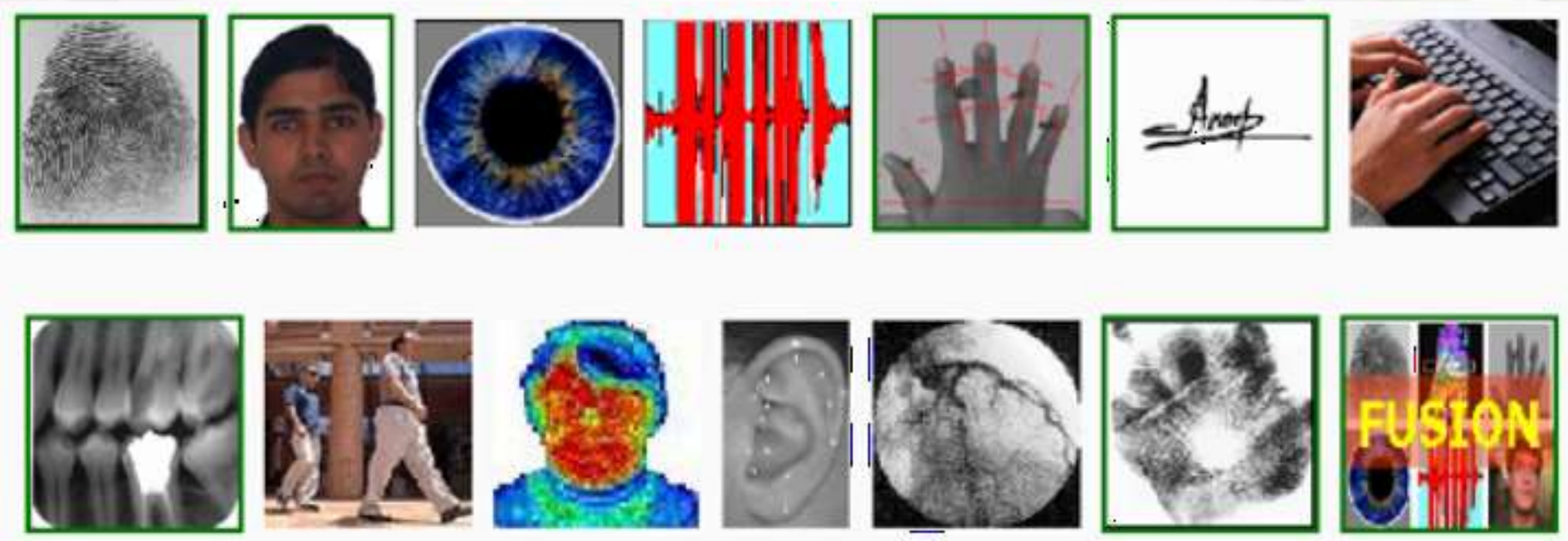
- **Text Processing**
  - Text Recognition (Printed / Handwritten Text Recognition, License Plate Recognition)
  - Office Automation (Machine Translation)
  - Bank Automation (Check Recognition)
  - Post Office Automation (Letter Sorting)





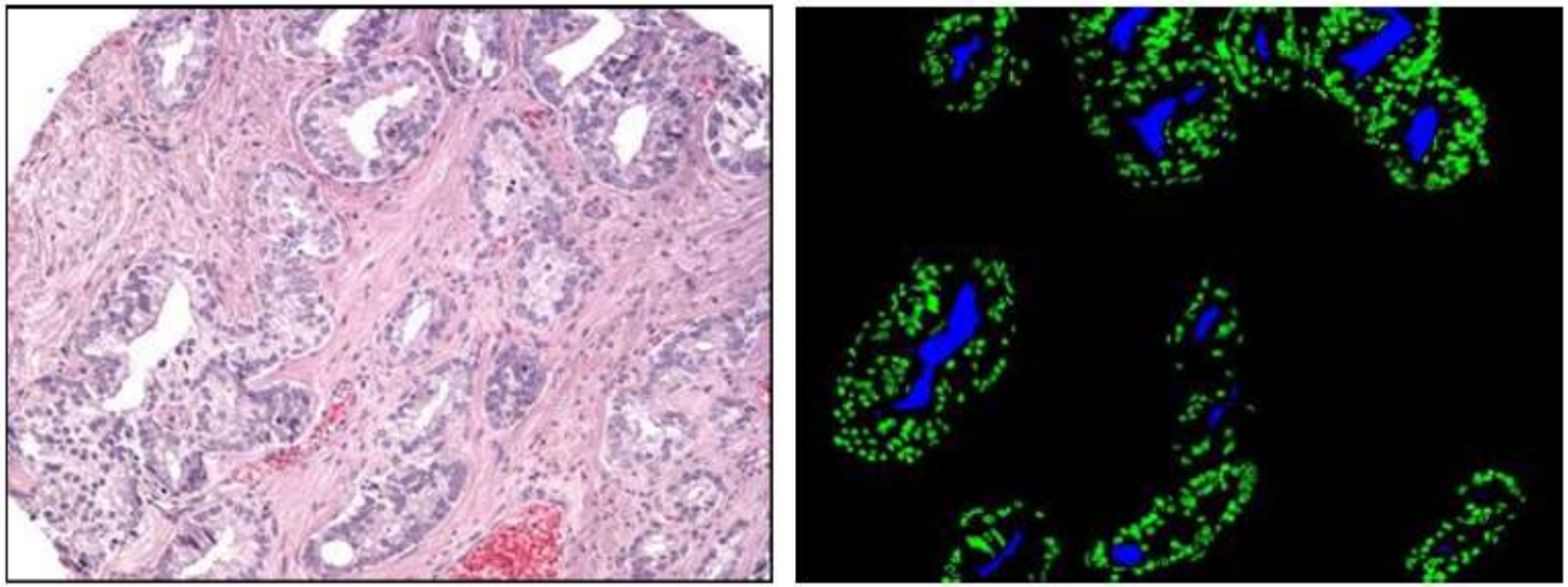
Appliation Areas of Artificial Neural Network

- **Biometrics**: Fingerprint Recognition, Face Recognition, Iris Recognition, Voice Recognition, Signature Recognition, Gait Recognition



Appliation Areas of Artificial Neural Network

- **Biomedicine**: Complete Blood Count (CBC), Chromosome Classification, Electrocardiography (ECG), Electroencephalogram (EEG), Plastic Surgery, Cancer Detection and Grading…



Cancer detection and grading using microscopic tissue data.

Appliation Areas of Artificial Neural Network

- **Remote Sensing**: Resource Investigation (Topographic and Geomorphologic Analysis, Lake Area Calculation), Map Recognition, Military Object Detection…

Appliation Areas of Artificial Neural Network

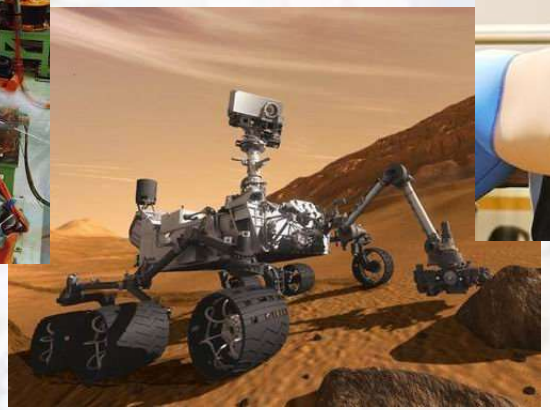
- Document Classification

Appliation Areas of Artificial Neural Network

- Forecast Decision

Appliation Areas of Artificial Neural Network

- **Robotics**
  - A robot is a programmable multi-functional operating device.
  - 4 stages of robotics research: telerobot, programmable robot, adaptive robot, intelligent robot

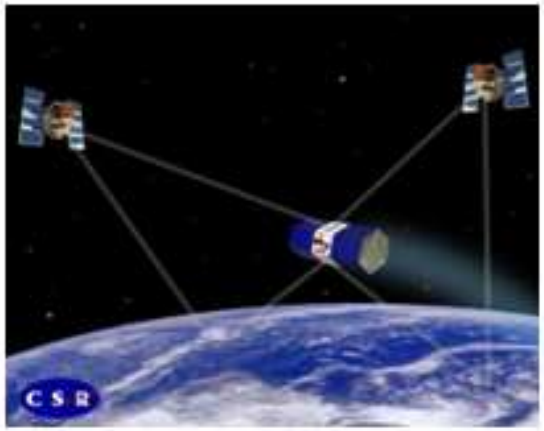




Appliation Areas of Artificial Neural Network

- Others…





Challenges towards Artificial Neural Network

- demand for large amount of data
- weak supervision of data label
- overfitting
- hyperparameter tuning
- opacity
- inflexibility
- …

Challenge #1: Demand for Large Amount of Data

In order to train a neural network that performs and generalizes well, we need a large training dataset, e.g. there are over 1 million images in ImageNet. To some extent, the performance of a neural network system is often limited to the size of the dataset.

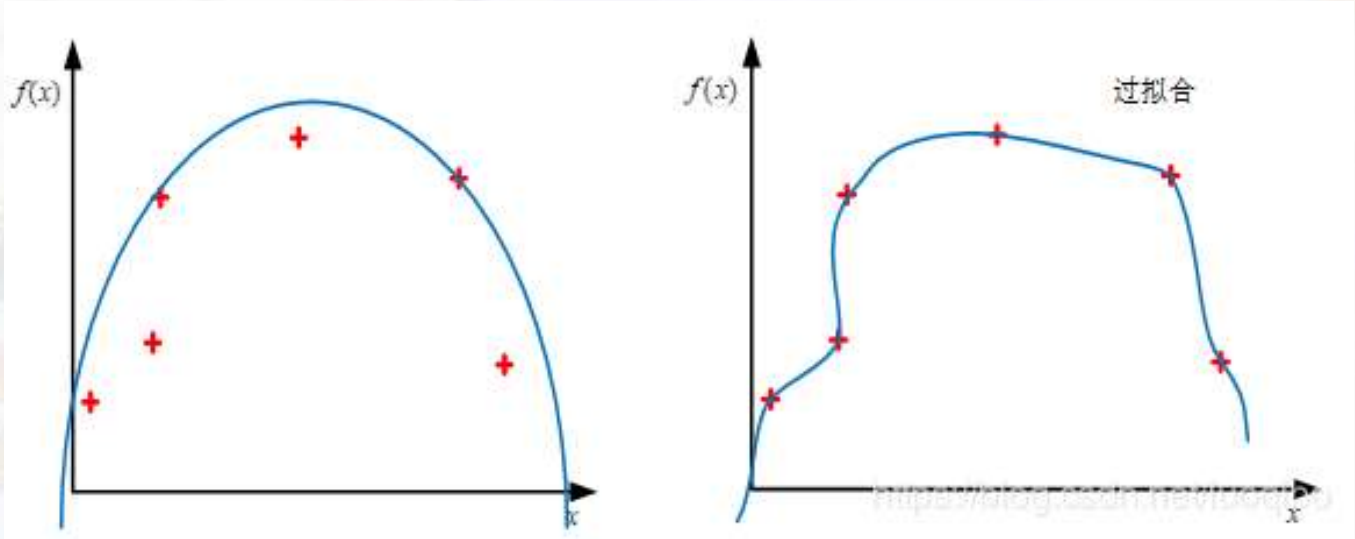
Challenge #2: Weak Supervision of Data Label

For most tasks in reality, due to the expensive manual data labeling, it is very challenging to obtain both accurate and complete labels for all samples in the dataset. There are various forms of weak supervision:

- **incomplete supervision**: Only parts of the training data are labeled;
- **inexact supervision**: The given labels are coarse-grained, e.g. we wish to construct an object localization neural network while the labels only tell us what categories exist in an image.
- **inaccucate supervision**: There exists noise in the given label, which is not always correct.

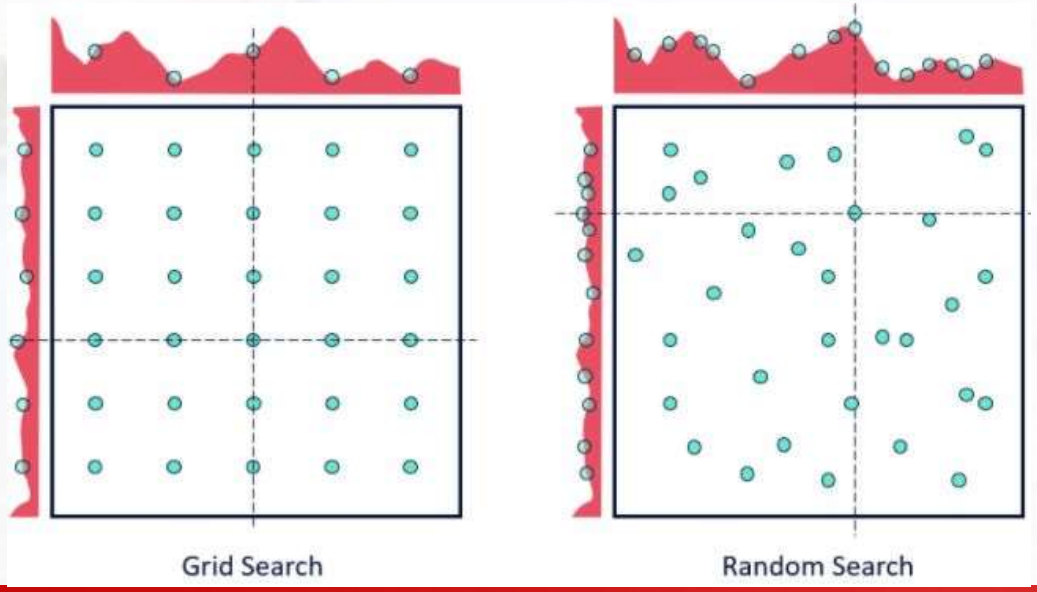
Challenge #3: Overfitting

A neural network is optimized using the collected training data, and hopes to have the same excellent performance on the unknown test data, thereby making applications in the real world. Improper training strategies will cause neural networks to only remember the training samples, instead of generalizing to new situations and samples.



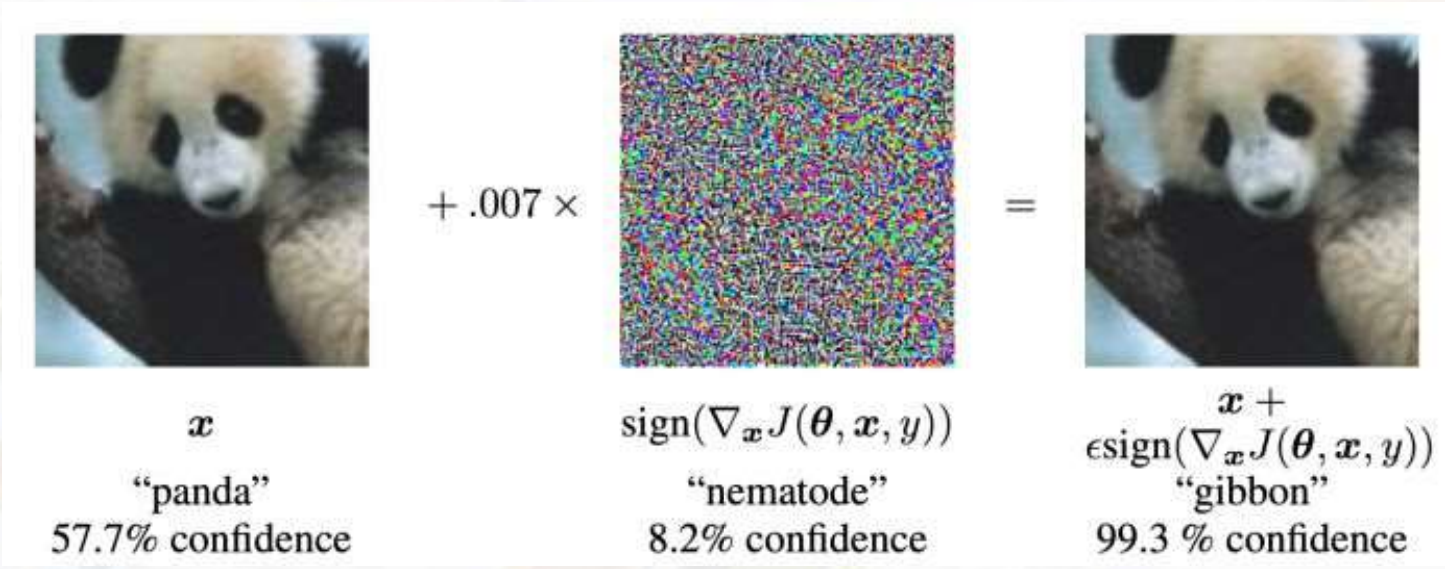
Challenge #4: Hyperparameter Tuning

Parameters in a neural network can be learned automatically during the training process, while hyperparameters need to be set manually before the training starts. Small tweaks towards hyperparameters can lead to large difference as for performance. Unfortunately, hyperparameters are decided based on researchers' historical experience usually.



Challenge #5: Opacity

There are millions of nodes of hidden-layers in modern neural networks. Essentially, A neural network is a black-box. Given an input, only the output can be observed, while the internal learning and reasoning process is basiccally invisible to humans. For those occasions where high reliability is required, such a model can lead to serious consequences.



Challenge #6: Inflexibility

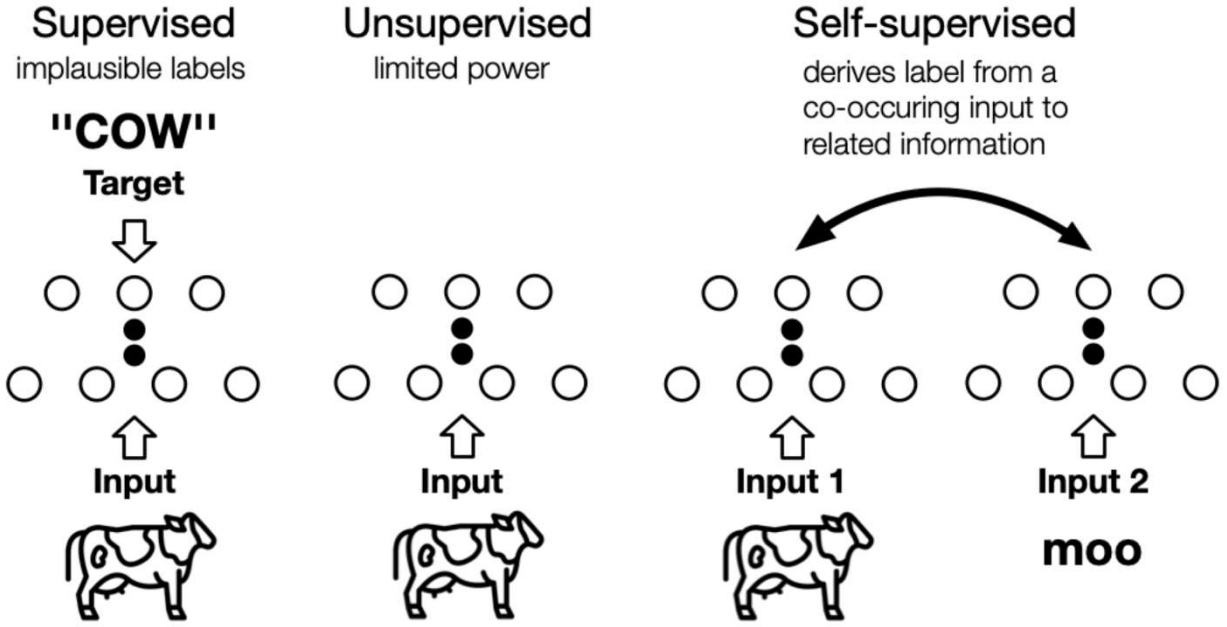
Currently, most neural network systems are specifically designed for a specific task. No matter how good they are at solving the original task, they usually can't be directly applied to a new task. After all, these neural networks are weak AI, which is so far away from strong AI.

AlphaGo is a typical weak AI, which can only handle the task of playing go.

Frontier #1: Self-Supervised Learning

- avoiding the demand for large amount of labeled data

Labeling the dataset requires lots of labor and time costs. By designing reasonable "auxiliary tasks", Self-Supervised Learning can learn discriminative features inside a sample from a large number of unlabeled samples, thereby replacing Fully Supervised Learning.



Frontier #1: Self-Supervised Learning

- avoiding the demand for large amount of labeled data

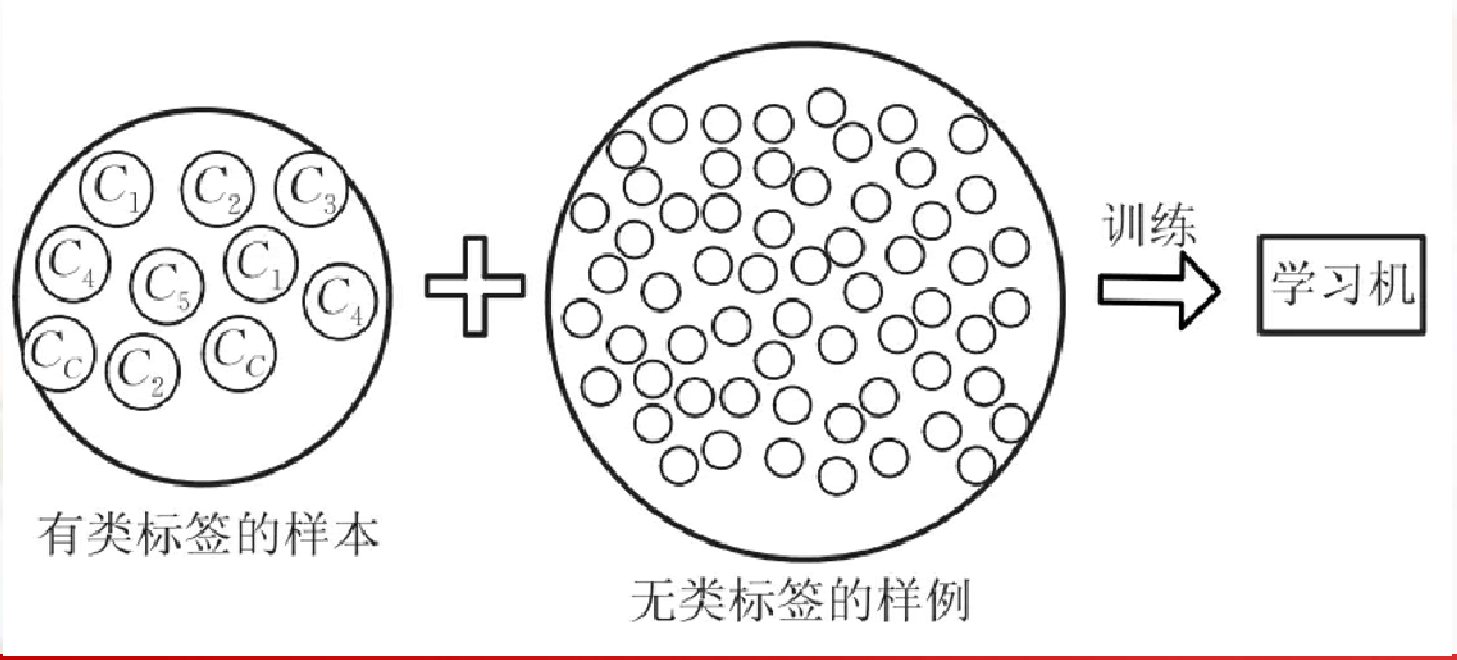
- "Pure" Reinforcement Learning (cherry)
  - The machine predicts a scalar reward given once in a while.
  - **A few bits for some samples**
- Supervised Learning (icing)
  - The machine predicts a category or a few numbers for each input
  - Predicting human-supplied data
  - **10→10,000 bits per sample**
- Unsupervised/Predictive Learning (cake)
  - The machine predicts any part of its input for any observed part.
  - Predicts future frames in videos
  - **Millions of bits per sample**

- (Yes, I know, this picture is slightly offensive to RL folks. But I'll make it up)

Frontier #2: Semi-Supervised Learning

- neural networks under incomplete supervision

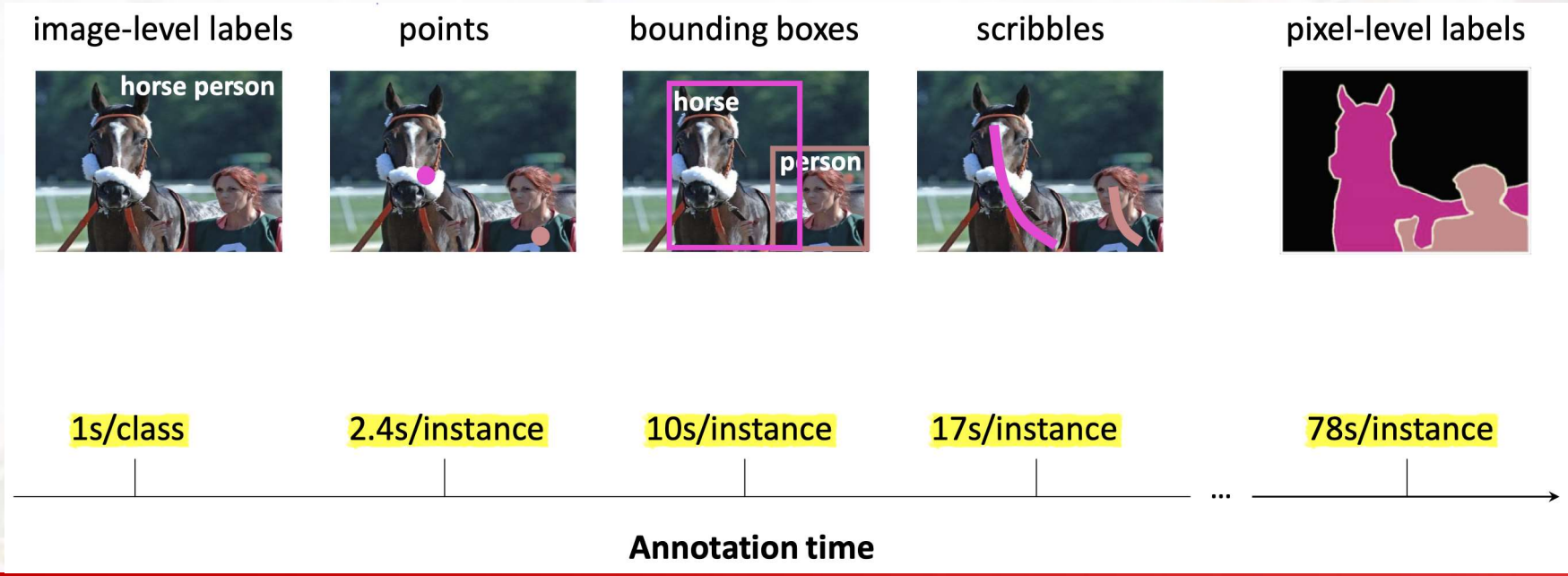
In most cases, there is not only a large amount of unlabeled data, but also a small amount of labeled data, resulting in Semi-Supervised Learning, which solves both the high labeling cost problem of Fully Supervised Learning and the inferior performance of Unsupervised Learning.



Frontier #3: Weakly Supervised Learning

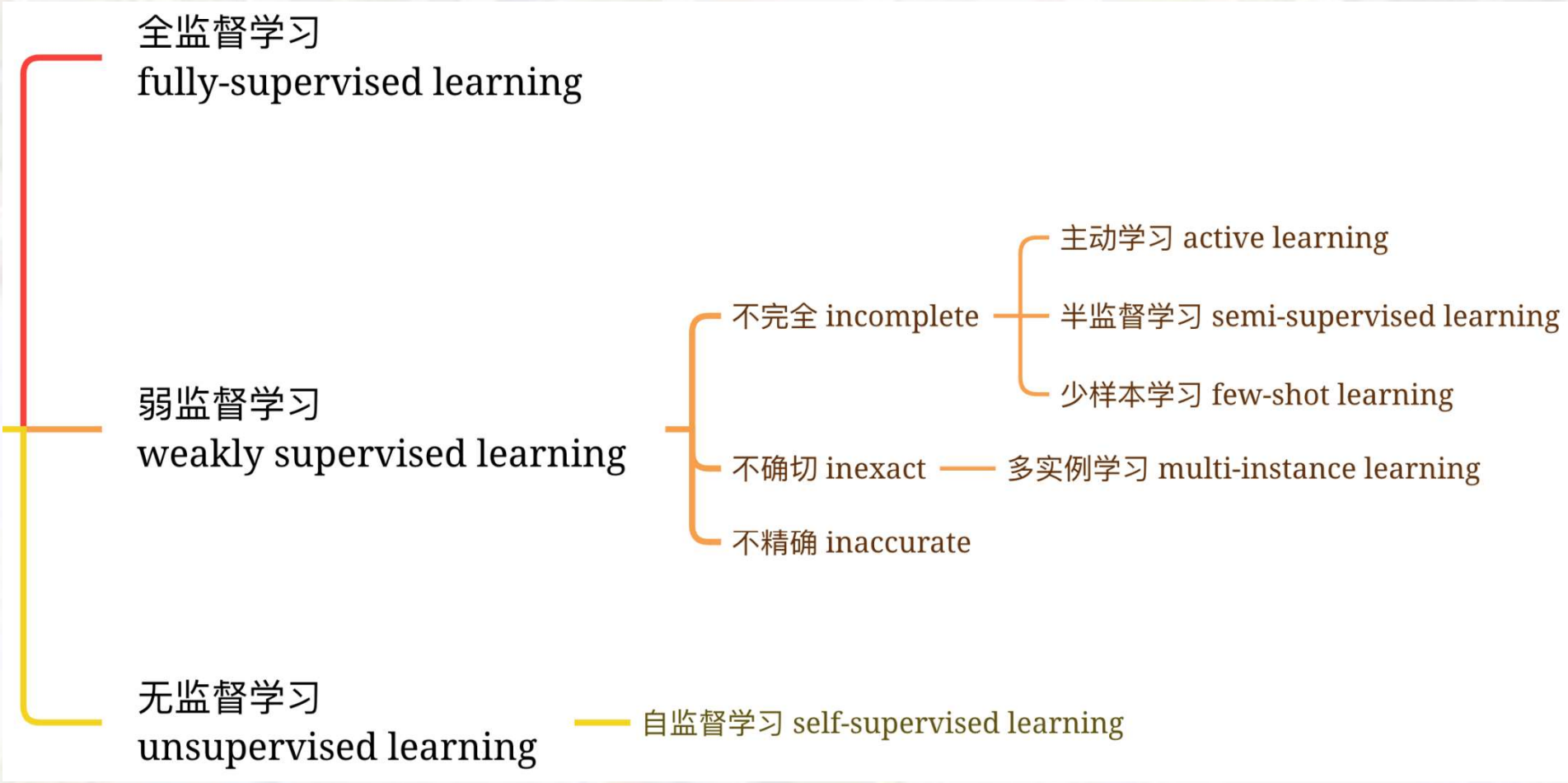
- neural networks under inexact supervision

The label granularity required for the task is often not aligned with the label granularity in the actual data. Weakly Supervised Learning can handle a variety of weak label forms, e.g. image-level labels, point labels, bounding box labels, scribble labels, pixel-level labels.



Frontier #3: Weakly Supervised Learning

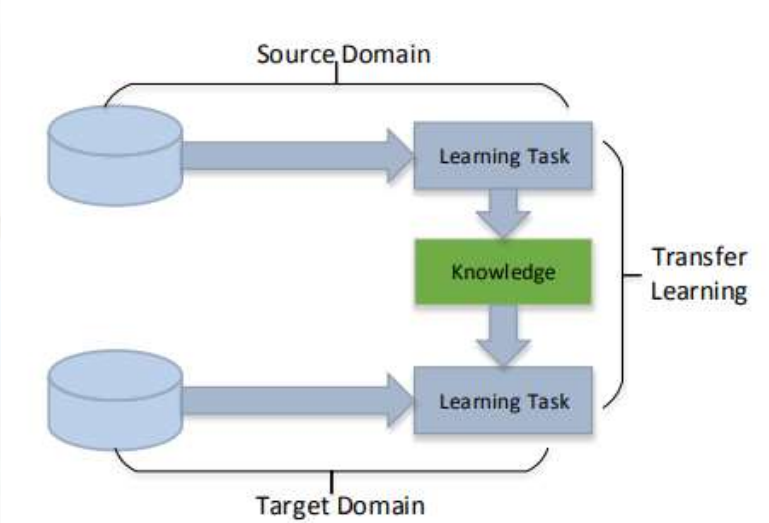
- neural networks under inexact supervision



Frontier #4: Transfer Learning and Domain Adaptation

- towards generalization of neural networks

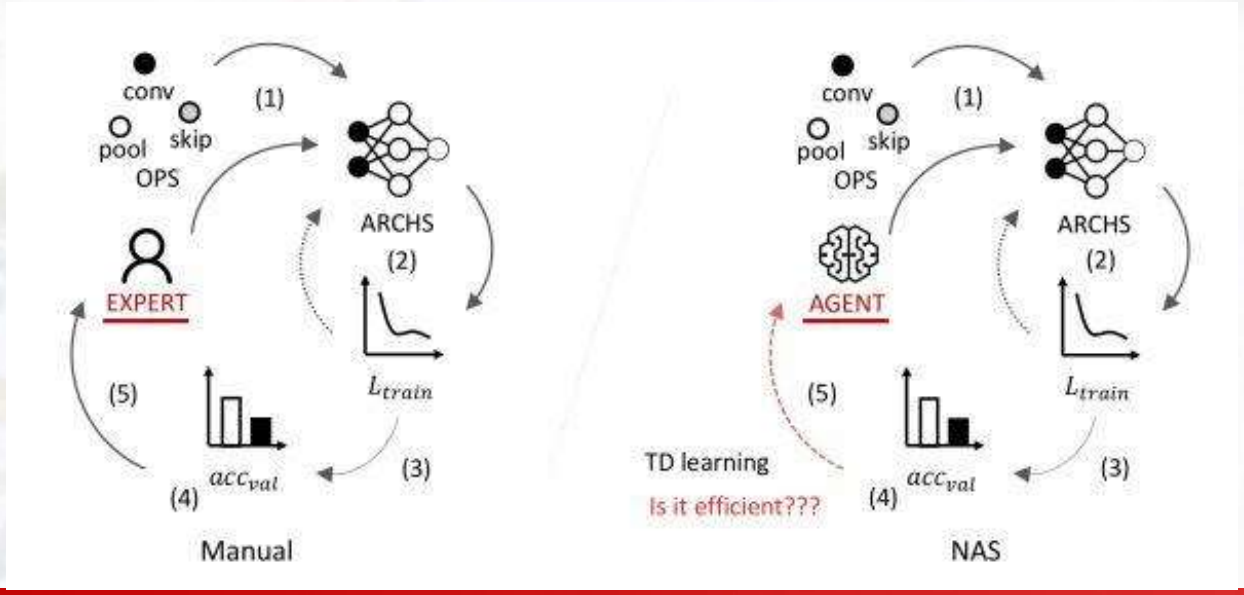
Transfer Learning, including Domain Adaptation, refers to transferring to new knowledge with the help of existing knowledge, which focuses on exploring the similarity between existing knowledge and new knowledge, thereby draw inferences from one case and improve the generalization ability across different domains.



Frontier #5: Neural Architecture Search

- learn to design a neural network automatically

Neural Architecture Search (NAS) takes the task of "designing a neural network" as the optimization objective of a neural network. It uses a set of candidate neural network structures, called "search space", to automatically search high-performance network structure through a certain strategy, sometimes even comparable to human experts.



Frontier #6: Interpretability

- understand how and why

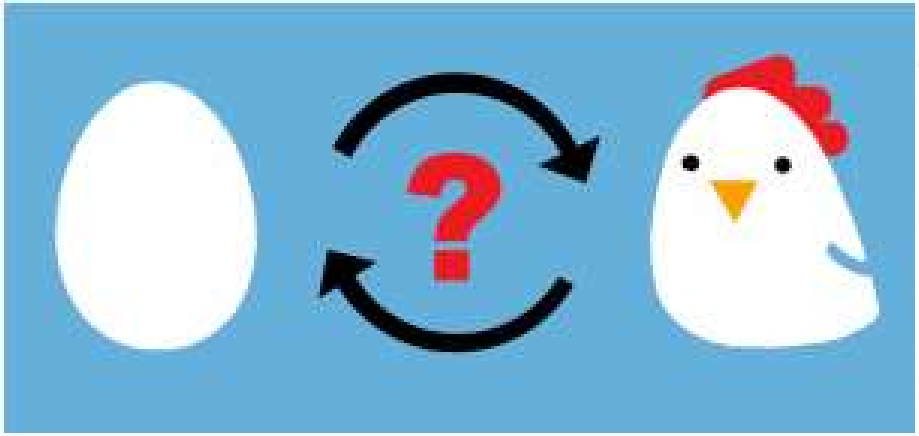
Neural Network Interpretability refers to describing "understanding of decision-making or prediction results towards a neural network", so that neural networks could not only tell us "what the result is", but also "why the result is obtained".

For image classification tasks, "class activation map" is a typical tool for interpretability, explaining which pixels are classified by the neural network.

Frontier #7: Causal Inference

- From result to cause

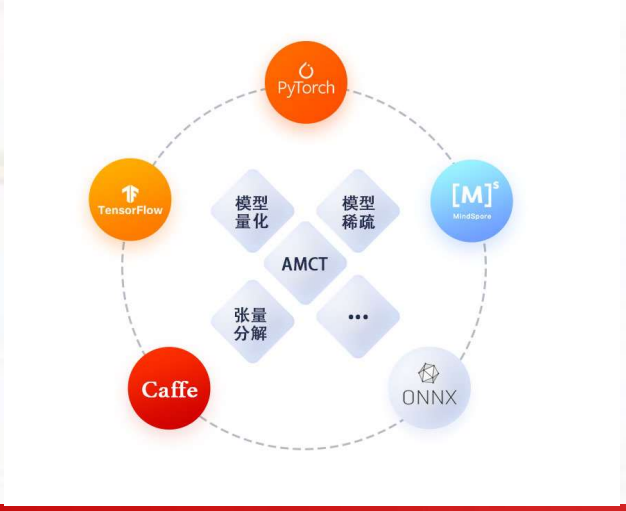
Limited by the distribution difference between the real world and the training data, we hope that neural networks can learn the essential reasons and laws to make predictions, rather than relying on superficial correlations. Causal Inference draws conclusions about causal relationships based on the conditions under which a result occurs, and has been closely integrated with deep learning methods in recent years.



Frontier #8: Lightweight Neural Network

- model compression and acceleration

Different from traditional servers, mobile devices have great challenges in deploying neural networks. The model must meet the conditions of small size, low computational complexity, and low power consumption. Lightweight Neural Network improves the performance of deployment by reducing the amount of parameters or computational complexity, e.g. network quantization, network pruning, and model distillation.



Frontier #9: Deep Learning framework

- integration of training, testing and deployment

Using an integrated deep learning framework to build the model can not only save the need for writing a lot of low-level codes, but also avoid the trouble of deployment and configuration. Nowadays, a large number of open-source deep learning frameworks, which can be used in production environments, are built by many domestic and foreign companies.

Frontier #9: Deep Learning framework

- integration of training, testing and deployment

- UC Berkeley: Caffe
- Google: TensorFlow, Keras
- Facebook: PyTorch
- Microsoft: CNTK
- Baidu: PaddlePaddle
- Huawei: MindSpore
- Megvii: MegEngine
- Tsinghua University: Jittor

Frontier #10: Transformer

- new structure of neural network

Different from CNN in the vision field and RNN in the text field, Transformer make use of a novel Self-Attention mechanism to model parallel relationships between different positions in a sequence, solving the long-range dependency problem in RNN.

Today, Transformer has shined in many fields such as vision (Vision Transformer, Swin Transformer), text (BERT, GPT-3), etc. Since it was proposed in 2017, Transformer has been cited more than 40,000 times. Obviously a trend of unifying neural network structures is caused by Transformer.


THANK
YOU
谢谢！