



Wenmin Wang

Principles of Machine Learning

The Three Perspectives

 Springer

Principles of Machine Learning

Part I Perspectives

1 Introduction

2 On Perspectives

1 Introduction

Contents

- 1.1 From “Queen” to “King”
- 1.2 Related Fields of Machine Learning
- 1.3 History of Machine Learning
- 1.4 Applications of Machine Learning
- 1.5 The Too-Small World

Machine learning used to be the Queen.

Machine learning is now the King.



http://www.computervisionblog.com/2015_06_01_archive.html

Origins of Machine Learning

- In 1946, ENIAC (Electronic Numerical Integrator and Computer), the first electronic general-purpose digital computer was officially released.
- In 1950, Alan Turing published “Computing Machinery and Intelligence” in *Mind* journal, and proposed learning machine in the article.
- In 1956, the famous Dartmouth meeting was held, and marked the official birth of artificial intelligence (AI) as a research field.
- In 1959, Arthur Samuel published “Some Studies in Machine Learning Using the Game of Checkers”, popularized the term “machine learning”.



Arthur Samuel

Definitions of Machine Learning

Who	When	Definition
Arthur Samuel (American pioneer of ML)	1959	The field of study that gives computers the ability to learn without being explicitly programmed.
Herbert Simon (American computer scientist)	1983	Learning is any process by which a system improves performance from experience.
Thomas Mitchell (Prof., Carnegie Mellon Univ)	1997	A computer program is said to learn from experience E with respect to some class of tasks T and performance measure P , if its performance at tasks in T , as measured by P , improves with experience E .
Ethem Alpaydin (Prof., Özyegin Univ, Turkish)	2004	Programming computers to optimize a performance criterion using example data or past experience.
Mehryar Mohri, et al. (Prof., New York Univ)	2012	The computational methods using experience to improve performance or to make accurate predictions.
Peter Flach (Prof., Univ of Bristol)	2012	Machine learning is about using the right features to build the right models that achieve the right tasks.

Definition of Machine Learning

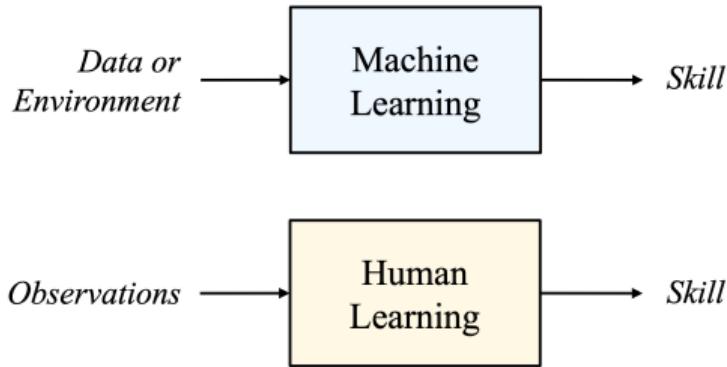
Definition: (Machine Learning)

Machine learning is a branch of artificial intelligence that studies and builds models that can learn from data or the environment to make predictions about unknown data or make decisions based on environmental information.

What's different in above definition

- It also defines the relationship between machine learning and AI.
- It also considers the factor of learning from the environment.

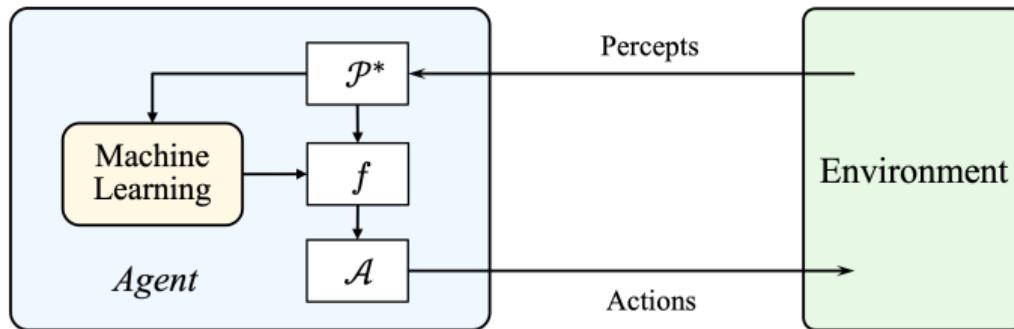
Machine Learning vs. Human Learning



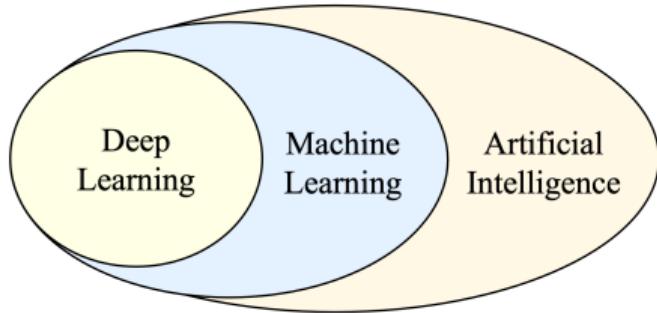
- Machine learning requires ready-made data or information feedback from the environment, its skills mainly manifest in prediction or decision making.
- Human can use their senses to observe the external world, the range covered by human skills is much wider.

Artificial Intelligence

- Artificial intelligence is the study of how to build an intelligent agent that can perceive the external environment and take corresponding actions to achieve a certain goal.
- Machine learning is how to build a model that can learn from data or the environment and use it to make predictions or decisions.



Deep Learning



- An artificial neural network with many hidden layers is called a deep neural network.
- The machine learning method using deep neural networks is called deep learning.
- Deep learning is a subfield of machine learning.

Statistical Learning

- Statistical learning theory is a framework for machine learning drawing from the fields of statistics and functional analysis.

The screenshot shows the Wikipedia article page for "Statistical learning theory". The page title is "Statistical learning theory". The left sidebar contains a "Contents" section with links to "Introduction", "Formal description", "Loss functions", "Regression", and "Classification". The main content area starts with a summary: "From Wikipedia, the free encyclopedia". Below the summary is a paragraph about the article's scope: "This article is about statistical learning in machine learning. For its use in psychology, see [Statistical learning in language acquisition](#). See also: [Computational learning theory](#)". A note at the bottom states: "Statistical learning theory is a framework for machine learning drawing from the fields of [statistics](#) and [functional analysis](#).^{[1][2][3]} Statistical learning theory deals with the". To the right of this note is a box stating "Part of a series on [Machine learning](#)". The top navigation bar includes a search bar with the placeholder "Search Wikipedia" and a "Search" button. There are also links for "Article", "Talk", "Read", "Edit", "View history", and "Tools". A language selector indicates "9 languages".

https://en.wikipedia.org/wiki/Statistical_learning_theory

Pattern Recognition

- It focuses on how to recognize patterns and regularities in input data.
- Traditional pattern recognition originates from statistics and engineering.
- Modern pattern recognition has gradually turned to processing patterns in data based on machine learning.

Data patterns + machine learning

- It has extended to biometric recognition and developed into biometrics technology, includes two categories:
 - ▶ physiological characteristics recognition,
 - ▶ behavioral characteristics recognition.

Data Mining

- It is the process of extracting and discovering meaningful patterns and useful information in large datasets.
- It belongs to the cross-research field of database systems and machine learning.

Database system + machine learning

- Data mining focuses on extracting and discovering meaningful patterns and useful information in data.
- Machine learning focuses on extracting certain features from data and using those features for making predictions.

Other Related Fields

- Computer vision
- Natural language processing
- Computer speech
- Robotics
- Bioinformatics
- Smart medical diagnosis
- etc.

About History of Machine Learning

Machine learning has developed in the process of constantly exploring how to build effective machine learning models, and it has also developed in the situation where the demand for machine learning in many fields is constantly increasing, so many kinds of algorithms for machine learning have been produced.

Through refinement and in-depth exploration, it can be found that machine learning mainly has five major development trajectories, those are:

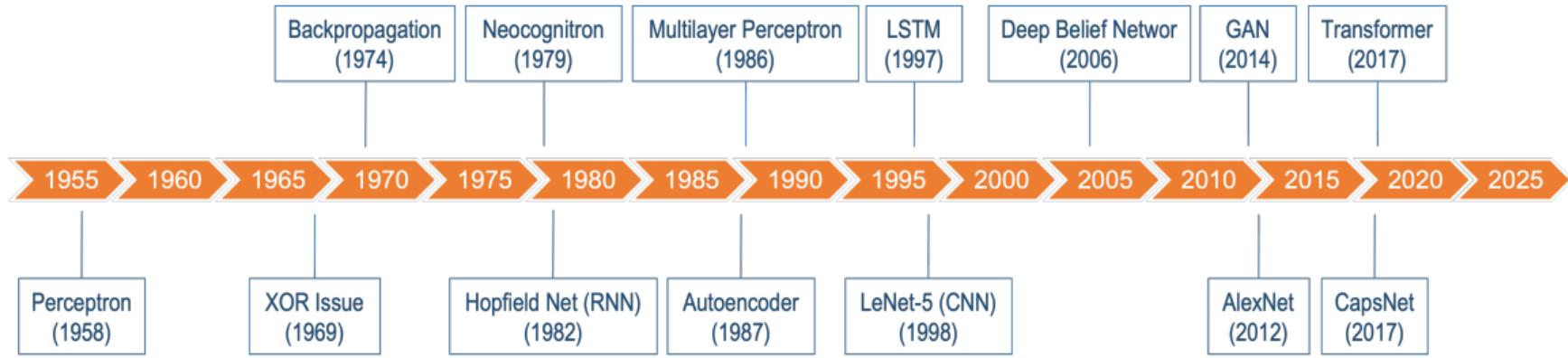
- artificial neural networks,
- decision trees,
- boosting algorithms,
- support vector machines,
- reinforcement learning.

Artificial Neural Networks

What are Artificial Neural Networks

An artificial neural network is a computing system built inspired by biological neural network in the brain, and it is also an implementation approach to machine learning.

Artificial Neural Networks



The timeline of the development of artificial neural networks.

Artificial Neural Networks

- In 1943, Warren McCulloch and Walter Pitts proposed threshold logic unit (TLU), the first artificial neuron in history.
- In 1958, Frank Rosenblatt invented Perceptron, a pattern recognition algorithm for binary classifier with only single-layer.
- In 1969, Marvin Minsky and Seymour Papert published Perceptrons, pointed out that perceptron can not for solving XOR problems.
- In 1970, Seppo Linnainmaa first described automatic differentiation (AD) method, that was the prototype of backpropagation (BP) algorithm.
- In 1974, Paul Werbos proposed possibility of using Seppo Linnainmaa's BP algorithm for artificial neural networks.
- In 1979, Kunihiko Fukushima et al. published Neocognitron, a multilayer network composed of neuron-like cells for visual pattern recognition.

Artificial Neural Networks

- In 1982, John Hopfield proposed content-addressable memory system, became the foundational literature for recurrent neural networks (RNNs).
- In 1985, Geoffrey Hinton et al. invented Boltzmann machine, another symmetrically connected network with neuron-like units.
- In 1986, David Rumelhart, Geoffrey Hinton and Ronald Williams published using BP algorithm to train multilayer perceptron (MLP).
- In 1987, Yann LeCun proposed the concept of autoencoders (AE).
- In 1997, Sepp Hochreiter and Jurgen Schmidhuber proposed long short-term memory (LSTM).
- In 1998, Yann LeCun et al. published LeNet-5, a 7-layer convolutional neural network (CNN). In 2006, Geoffrey Hinton et al. published deep belief network (DBN).

Artificial Neural Networks

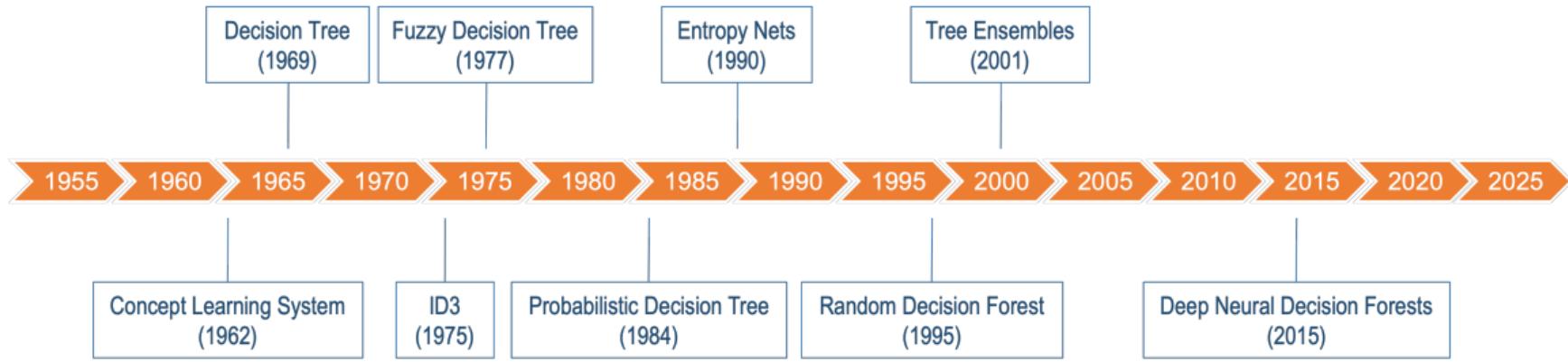
- In 2011, Jeff Dean, Greg Corrado and Andrew Ng formed Google Brain team to build a large-scale deep learning system.
- In 2012, Alex Krizhevsky et al. designed AlexNet that achieved excellent results in ImageNet Large Scale Visual Recognition Challenge (ILSVRC).
- In 2014, Ian Goodfellow et al. proposed the generative adversarial network (GAN), a novel unsupervised deep learning algorithm.
- In 2017, Ashish Vaswani et al. published Transformer, a deep learning architecture that relies on the parallel attention mechanism, which has led to the development of large language models (LLMs).
- In 2017, Jeffrey Hinton proposed capsule network (CapsNet) based on the theory of cortical mini-columns in neuroscience.

Decision Trees

What are Decision Trees

A decision trees breaks down the complex decision problems into a tree-like structure of all decision options and subsequent chance events related to specific options.

Decision Trees



The main development history of decision trees.

Decision Trees

- In 1962, Earl Hunt et al. proposed concept learning system (CLS), for constructing decision tree that minimizes cost of classifying objects.
- In 1969, Patrick Winston published “A Heuristic Program That Constructs Decision Trees”.
- In 1975, John Quinlan published the classification algorithm, ID3 (Iterative Dichotomiser 3), for generating decision trees from fixed samples.
- In 1984, Richard Casey and George Nagy published “Decision Tree Design Using a Probabilistic Model” for optical character recognition.
- In 1990, Ishwar Sethi published entropy networks, that converts decision trees into neural networks.

Decision Trees

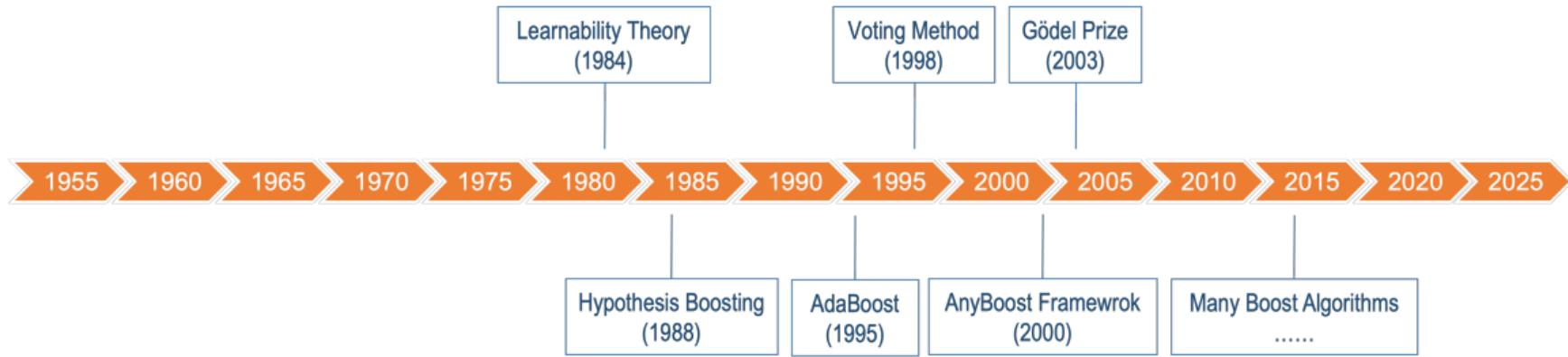
- In 1995, Tin Kam Ho developed random decision forest, also known as random forest, based on decision trees.
- In 2001, Leo Breiman discussed that random forests belong to ensemble of multiple trees and added bagging method and so on into random forests.
- In 2015, Peter Kotschieder et al., published “Deep Neural Decision Forests”, proposing a classification method that integrates deep convolutional neural networks and decision trees.

Boosting Methods

What are Boosting Methods

The basic idea of the boosting methods in machine learning is to integrate several weak learners into a strong learner.

Boosting Methods



The main development trajectory of the boosting methods.

Boosting Methods

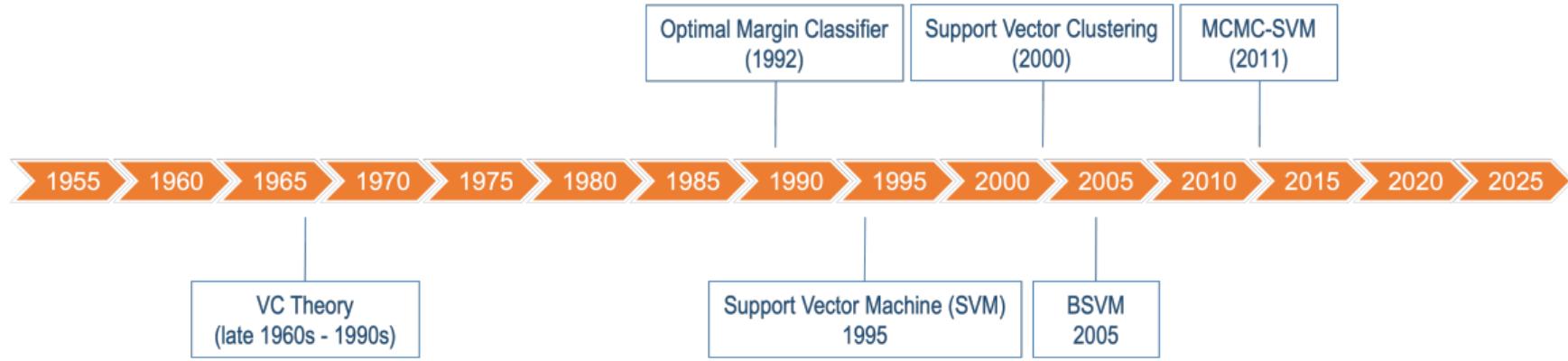
- Originated from an extension of theory of learnability: strong learnability and weak learnability.
- In 1988, Michael Kearns gave a formal definition of strong learnability and weak learnability, and proposed hypothesis boosting problem.
- In 1990, Robert Schapire proved hypothesis boosting problem, and proposed a method to boost weak learning algorithms into any strong learner.
- In 1995, AdaBoost was proposed by Yoav Freund and Robert Schapire. It can combine many types of weak classifiers, into a strong classifier.
- In 2000, Llew Mason et al. proposed a framework, AnyBoost, and proved that many voting methods are special cases of AnyBoost.

Support Vector Machines

What are Support Vector Machines

Support vector machines (SVMs) originated from VC Theory, and inspired by the method for maximizing the decision boundary of classification.

Support Vector Machines



The main development history of support vector machines.

Support Vector Machines

- During 1960-1990, Vladimir Vapnik and Alexey Chervonenkis proposed VC Theory to explain learning process from a statistical perspective.
- In 1992, Bernhard Boser et al. published “A Training Algorithm for Optimal Margin Classifiers” for maximizing decision boundary of classification.
- In 1995, Corinna Cortes and Vladimir Vapnik published “Support-Vector Networks”, which is also called support vector machines.
- In 1996, Drucker et al. applied SVM to regression task, and published “Support Vector Regression Machines”.
- In 1999, Erin Bredensteiner and Kristin Bennett proposed a method to implement multi-classification support vector machines.

Support Vector Machines

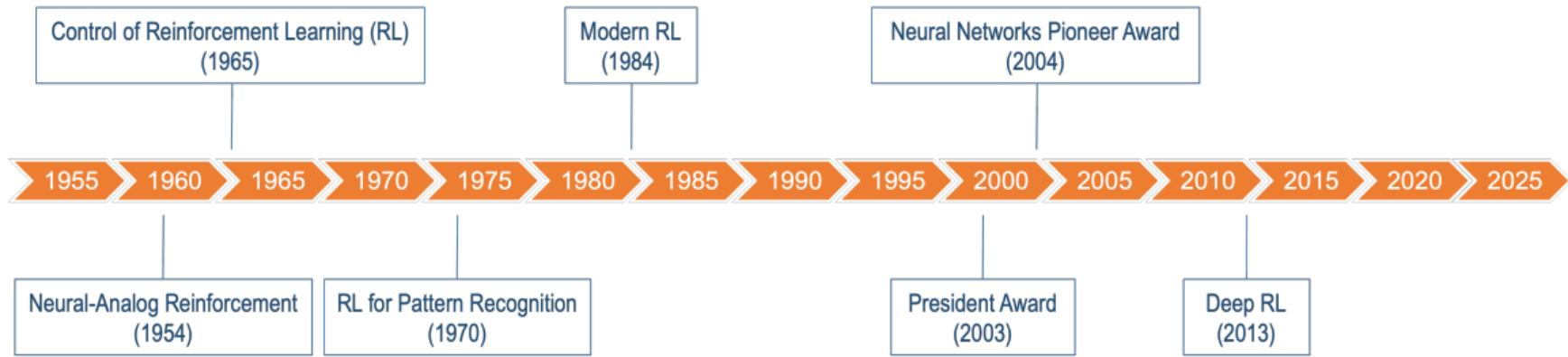
- In 2000, Asa Ben-Hur et al. proposed a kernel method for data clustering, called support vector clustering (SVC) algorithm.
- In 2005, Mallick et al. proposed Bayesian support vector machine (BSVM) based on Bayesian theorem, for classifying tumors.
- In 2011, Nicholas Polson and Steven Scotty proposed a latent variable representation of regularized SVMs that enables expectation-maximization (EM), expectation-conditional maximization (ECM), or Markov chain Monte Carlo (MCMC) algorithms to provide parameter estimates.

Reinforcement Learning

What is Reinforcement Learning

Reinforcement learning (RL) is the approach to conduct how intelligent agents improve their behavior according to the state and rewards of the environment.

Reinforcement Learning



The main development progress of reinforcement learning.

Reinforcement Learning

- In 1954, Marvin Minsky studied the theory of neural-analog reinforcement systems and its application in his doctoral thesis.
- In 1965, M. D. Waltz and King-sun Fu published reinforcement learning control systems for non-linear and non-stationary factory equipment.
- In 1970, Jerry Mendel and Robert McLaren published reinforcement learning control and pattern recognition systems.
- In 1984, Richard Sutton published his doctoral thesis “Temporal Credit Assignment in Reinforcement Learning”.

Sutton's PhD thesis advisor was professor Andrew Barto of the University of Massachusetts Amherst.

Reinforcement Learning

- In 2003, International Neural Network Society awarded Richard Sutton the President Award, due to his contributions to reinforcement learning.
- In 2004, IEEE Neural Networks Society awarded Andrew Barto the Neural Networks Pioneer Award, to his contributions to reinforcement learning.
- In 2013, Mnih et al. combined Q -learning algorithm in reinforcement learning with CNN, called deep reinforcement learning (deep RL).
- Since 2015, AlphaGo Fan, AlphaGo Lee, AlphaGo Master, and AlphaGo Zero are proposed, by combining deep RL with Monte Carlo tree search.
- In 2017, AlphaZero used an approach similar to AlphaGo Zero, can play three types of board games, Chess, Shogi and Go.

Computer Vision

- What it is: to make computers process images or videos at a higher level.
- Why it matters: The investigation at Harvard Business School, the proportions of information humans need to process daily: taste 1%, touch 1.5%, smell 3.5%, hearing 11%, and vision 83%.
- What tasks mainly:
 - image restoration,
 - event detection,
 - video tracking,
 - object segmentation,
 - object recognition,
 - pose estimation,
 - motion estimation,
 - scene modeling,
 - scene reconstruction,
 - 3D reconstruction,
 - image understanding,
 - visual generation.

Computer Speech

Including speech recognition and speech synthesis.

Speech Recognition

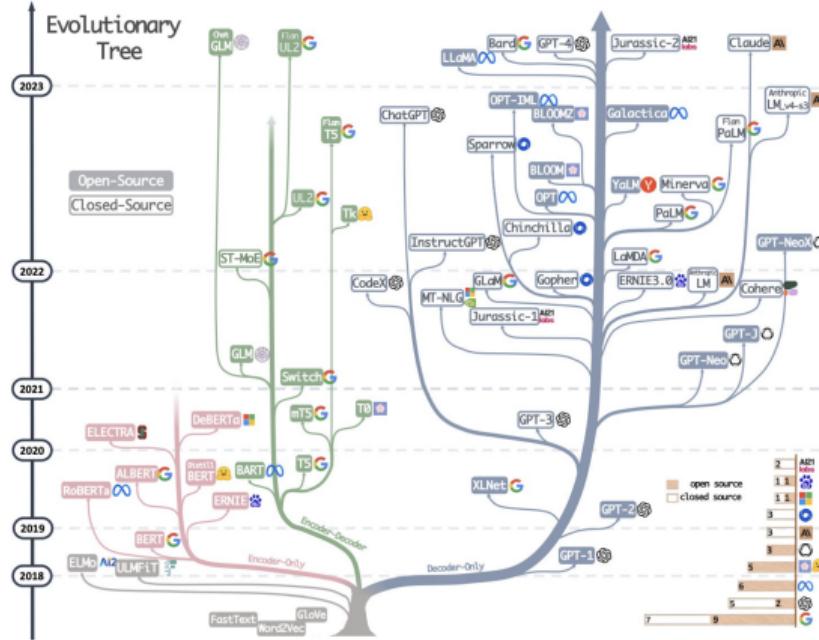
- To recognize speech and translate it into text.
- Apple's Siri, Microsoft's Cortana, Baidu's Duer, Amazon's Alexa.

Speech Synthesis

- To convert text into audio information and generate human speech.
- GPT-4o demo: <https://openai.com/index/hello-gpt-4o/>

Large Language Models

- OpenAI's GPT
- Google's Gemini
- Meta's LLaMa
- Baidu's Ernie
- Huawei's Pangu
- Anthropic's Claude
- Amazon's Olympus
- etc.



<https://www.interconnects.ai/p/llm-development-paths>

Data Mining and Knowledge Discovery

Data Mining

- What it is: to extract and discovery unknown patterns from large data sets.
- What tasks mainly:
 - ▶ obtaining new data record groups through cluster analysis,
 - ▶ discovering abnormal records through anomaly detection,
 - ▶ finding association rules through dependency relationships.

Knowledge Discovery

- What it is: to extract and discovery knowledge from large data sets.
- Closely related data mining in terms of methodology and terminology.

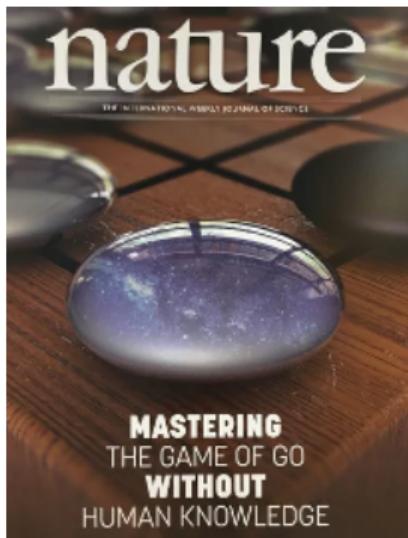
Board Games

Category	Typical game
Board Games	Checkers
	Chess
	Shogi
	Go
Card Games	Texas Hold'em
	Bridge
Tile games	Mahjong
Video Games	Black Myth: Wukong

The category of board games

Board Games

- Board games have simple game rules but vast game space, so the gameplay itself becomes very complex.
- DeepMind's AlphaZero has defeated all strongest AI game systems that master Chess, Shogi and Go.



<https://www.nature.com/articles/nature24270>



<https://www.science.org/doi/10.1126/science.aar6404>

Card Games

- Card games are imperfect information games that are different from perfect information games like board games.
- CMU and Facebook created an AI poker called Pluribus, defeated five top human players in six-player no-limit Texas Hold'em.

The screenshot shows a news article from the journal 'nature' (International Journal of Science). The header includes the 'nature' logo and a 'Subscribe' button. Below the header, the text 'NEWS • 11 JULY 2019' is displayed. The main headline reads: 'No limit: AI poker bot is first to beat professionals at multiplayer game'. A subtext below states: 'Triumph over five human opponents at Texas hold'em brings bots closer to solving complicated real-world problems.' The entire article is contained within a white rectangular box.

<https://www.nature.com/articles/d41586-019-02156-9>

The screenshot shows a news article from the journal 'Science'. The header includes the 'Science' logo and navigation links for 'Contents', 'News', 'Careers', and 'Journals'. A 'CALL FOR PAPERS' banner for 'Science Signaling' is visible. The main headline reads: 'Superhuman AI for multiplayer poker'. Below the headline, the authors are listed as 'Noam Brown^{1,2,*}, Tuomas Sandholm^{1,3,4,5,*}' with a note 'See all authors and affiliations'. At the bottom, the publication details are given: 'Science 11 Jul 2019; issue 2440; DOI: 10.1126/science.aay2400'. The entire article is contained within a white rectangular box.

<https://www.science.org/doi/10.1126/science.aay2400>

Other Computer Games

Tile-based Games

- The most difficult game in tile-based games is Mahjong, which is also the game of imperfect information and multiple players like card games.
- Microsoft Research Asia developed powerful Mahjong AI called Suphx, promoted to 10th level on the Mahjong platform *Tenhou*.

Video Games

- OpenAI Five is the AI system that plays the five-on-five video game Dota 2, defeated the previous TI8 International champion team “OG”.
- AlphaStar, a computer game program developed by DeepMind, defeated two human professional players in real-time strategy game StarCraft II.

Autonomous Technologies

Self-driving Vehicles

- Be also known as driverless cars.
- Waymo driverless cars: <https://waymo.com/>

Drones

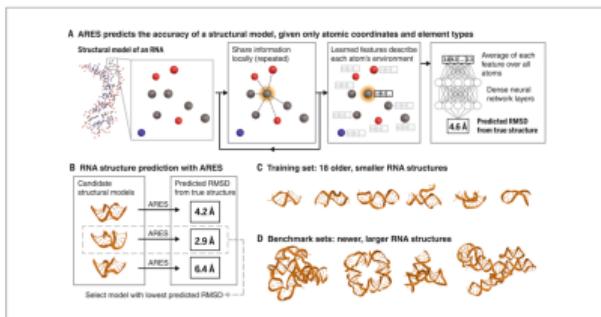
- Be unmanned aerial vehicles controlled remotely or autonomously.
- Flying cars are combination of drones and vehicles.

Robots

- Be the autonomous machines to perform a wide range of complex tasks.

Machine Learning in Medicine and Biology

- Medical diagnosis, drug synthesis.
- Protein structure determination: AlphaFold.
- Ribonucleic Acid (RNA) structure determination.



<https://www.science.org/doi/10.1126/science.abe5650>

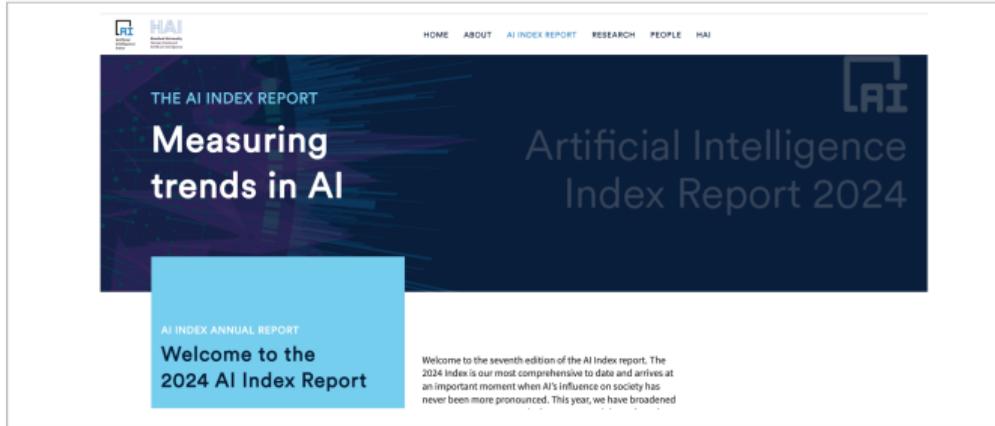


<https://www.science.org/doi/10.1126/science.abj8754>

Other Applications

- Sentiment analysis
- Fraud detection
- Spam filtering
- Recommender systems
- Time-series forecasting
- Cyber security
- etc.

Machine learning has entered their heyday



<https://aiindex.stanford.edu/report/>

The Too-small World of Artificial Intelligence

<https://towardsai.net/p/l/the-too-small-world-of-artificial-intelligence>

The Author's Comments

- Machine learning and artificial intelligence are sciences, and science cannot tolerate bubbles.
- Machine learning and artificial intelligence are also technologies, and only those that can be applied to real life and bring economic or social benefits can be called technologies.

2018 Turing Award



Yoshua Bengio



Geoffrey Hinton



Yann LeCun

Yoshua Bengio, Geoffrey Hinton, and Yann LeCun won the 2018 Turing Award: for conceptual and engineering breakthroughs that have made deep neural networks a critical component of computing.

<https://awards.acm.org/about/2018-turing/>

2024 Nobel Prize



John Hopfield



Geoffrey Hinton



Demis Hassabis



John Jumper

John Hopfield, and Geoffrey Hinton Won the Nobel Prize in Physics 2024:
for foundational discoveries and inventions that enable machine learning with ANNs.

Demis Hassabis, and John Jumper Won the Nobel Prize in Chemistry 2024:
“for protein structure prediction”.

<https://www.nobelprize.org/all-nobel-prizes-2024/>

Thank You