

Title: Deep reinforcement learning

URL: [https://en.wikipedia.org/wiki/Deep\\_reinforcement\\_learning](https://en.wikipedia.org/wiki/Deep_reinforcement_learning)

PageID: 60105148

Categories: Category:Deep learning, Category:Machine learning algorithms, Category:Reinforcement learning

Source: Wikipedia (CC BY-SA 4.0).

-----

Supervised learning

Unsupervised learning

Semi-supervised learning

Self-supervised learning

Reinforcement learning

Meta-learning

Online learning

Batch learning

Curriculum learning

Rule-based learning

Neuro-symbolic AI

Neuromorphic engineering

Quantum machine learning

Classification

Generative modeling

Regression

Clustering

Dimensionality reduction

Density estimation

Anomaly detection

Data cleaning

AutoML

Association rules

Semantic analysis

Structured prediction

Feature engineering

Feature learning

Learning to rank

Grammar induction

Ontology learning

Multimodal learning

Apprenticeship learning

Decision trees

Ensembles Bagging Boosting Random forest

Bagging

Boosting

Random forest

k -NN

Linear regression

Naive Bayes

Artificial neural networks

Logistic regression

Perceptron

Relevance vector machine (RVM)

Support vector machine (SVM)

BIRCH

CURE

Hierarchical

k -means

Fuzzy

Expectation–maximization (EM)

DBSCAN

OPTICS

Mean shift

Factor analysis

CCA

ICA

LDA

NMF

PCA

PGD

t-SNE

SDL

Graphical models Bayes net Conditional random field Hidden Markov

Bayes net

Conditional random field

Hidden Markov

RANSAC

k -NN

Local outlier factor  
Isolation forest  
Autoencoder  
Deep learning  
Feedforward neural network  
Recurrent neural network LSTM GRU ESN reservoir computing  
LSTM  
GRU  
ESN  
reservoir computing  
Boltzmann machine Restricted  
Restricted  
GAN  
Diffusion model  
SOM  
Convolutional neural network U-Net LeNet AlexNet DeepDream  
U-Net  
LeNet  
AlexNet  
DeepDream  
Neural field Neural radiance field Physics-informed neural networks  
Neural radiance field  
Physics-informed neural networks  
Transformer Vision  
Vision  
Mamba  
Spiking neural network  
Memtransistor  
Electrochemical RAM (ECRAM)  
Q-learning  
Policy gradient  
SARSA  
Temporal difference (TD)  
Multi-agent Self-play  
Self-play  
Active learning  
Crowdsourcing  
Human-in-the-loop

Mechanistic interpretability

RLHF

Coefficient of determination

Confusion matrix

Learning curve

ROC curve

Kernel machines

Bias–variance tradeoff

Computational learning theory

Empirical risk minimization

Occam learning

PAC learning

Statistical learning

VC theory

Topological deep learning

AAAI

ECML PKDD

NeurIPS

ICML

ICLR

IJCAI

ML

JMLR

Glossary of artificial intelligence

List of datasets for machine-learning research List of datasets in computer vision and image processing

List of datasets in computer vision and image processing

Outline of machine learning

v

t

e

Deep reinforcement learning ( deep RL ) is a subfield of machine learning that combines reinforcement learning (RL) and deep learning . RL considers the problem of a computational agent learning to make decisions by trial and error. Deep RL incorporates deep learning into the solution, allowing agents to make decisions from unstructured input data without manual engineering of the state space . Deep RL algorithms are able to take in very large inputs (e.g. every pixel rendered to the screen in a video game) and decide what actions to perform to optimize an objective (e.g. maximizing the game score). Deep reinforcement learning has been used for a diverse set of applications including but not limited to robotics , video games , natural language processing , computer vision , [ 1 ] education, transportation, finance and healthcare . [ 2 ]

Overview

## Deep learning

Deep learning is a form of machine learning that transforms a set of inputs into a set of outputs via an artificial neural network . Deep learning methods, often using supervised learning with labeled datasets, have been shown to solve tasks that involve handling complex, high-dimensional raw input data (such as images) with less manual feature engineering than prior methods, enabling significant progress in several fields including computer vision and natural language processing . In the past decade, deep RL has achieved remarkable results on a range of problems, from single and multiplayer games such as Go , Atari Games , and Dota 2 to robotics. [ 3 ]

## Reinforcement learning

Reinforcement learning is a process in which an agent learns to make decisions through trial and error. This problem is often modeled mathematically as a Markov decision process (MDP), where an agent at every timestep is in a state  $s$  , takes action  $a$  , receives a scalar reward and transitions to the next state  $s'$  according to environment dynamics  $p(s'|s,a)$  . The agent attempts to learn a policy  $\pi(a|s)$  , or map from observations to actions, in order to maximize its returns (expected sum of rewards). In reinforcement learning (as opposed to optimal control ) the algorithm only has access to the dynamics  $p(s'|s,a)$  through sampling.

## Deep reinforcement learning

In many practical decision-making problems, the states  $s$  of the MDP are high-dimensional (e.g., images from a camera or the raw sensor stream from a robot) and cannot be solved by traditional RL algorithms. Deep reinforcement learning algorithms incorporate deep learning to solve such MDPs, often representing the policy  $\pi(a|s)$  or other learned functions as a neural network and developing specialized algorithms that perform well in this setting.

## History

Along with rising interest in neural networks beginning in the mid 1980s, interest grew in deep reinforcement learning, where a neural network is used in reinforcement learning to represent policies or value functions. Because in such a system, the entire decision making process from sensors to motors in a robot or agent involves a single neural network , it is also sometimes called end-to-end reinforcement learning. [ 4 ] One of the first successful applications of reinforcement learning with neural networks was TD-Gammon , a computer program developed in 1992 for playing backgammon . [ 5 ] Four inputs were used for the number of pieces of a given color at a given location on the board, totaling 198 input signals. With zero knowledge built in, the network learned to play the game at an intermediate level by self-play and TD( $\lambda$ ) .

Seminal textbooks by Sutton and Barto on reinforcement learning, [ 6 ] Bertsekas and Tsitiklis on neuro-dynamic programming, [ 7 ] and others [ 8 ] advanced knowledge and interest in the field.

Katsunari Shibata's group showed that various functions emerge in this framework, [ 9 ] [ 10 ] [ 11 ] including image recognition, color constancy, sensor motion (active recognition), hand-eye coordination and hand reaching movement, explanation of brain activities, knowledge transfer, memory, [ 12 ] selective attention, prediction, and exploration. [ 10 ] [ 13 ]

Starting around 2012, the so-called deep learning revolution led to an increased interest in using deep neural networks as function approximators across a variety of domains. This led to a renewed interest in researchers using deep neural networks to learn the policy, value, and/or Q functions present in existing reinforcement learning algorithms.

Beginning around 2013, DeepMind showed impressive learning results using deep RL to play Atari video games. [ 14 ] [ 15 ] The computer player a neural network trained using a deep RL algorithm, a deep version of Q-learning they termed deep Q-networks (DQN), with the game score as the reward. They used a deep convolutional neural network to process 4 frames RGB pixels (84x84) as inputs. All 49 games were learned using the same network architecture and with minimal prior knowledge, outperforming competing methods on almost all the games and performing at a level comparable or superior to a professional human game tester. [ 15 ]

Deep reinforcement learning reached another milestone in 2015 when AlphaGo , [ 16 ] a computer program trained with deep RL to play Go , became the first computer Go program to beat a human professional Go player without handicap on a full-sized 19×19 board.

In a subsequent project in 2017, AlphaZero improved performance on Go while also demonstrating they could use the same algorithm to learn to play chess and shogi at a level competitive or superior to existing computer programs for those games, and again improved in 2019 with MuZero . [ 17 ] Separately, another milestone was achieved by researchers from Carnegie Mellon University in 2019 developing Pluribus , a computer program to play poker that was the first to beat professionals at multiplayer games of no-limit Texas hold 'em . OpenAI Five , a program for playing five-on-five Dota 2 beat the previous world champions in a demonstration match in 2019.

Deep reinforcement learning has also been applied to many domains beyond games. In robotics, it has been used to let robots perform simple household tasks [ 18 ] and solve a Rubik's cube with a robot hand. [ 19 ] [ 20 ] Deep RL has also found sustainability applications, used to reduce energy consumption at data centers. [ 21 ] Deep RL for autonomous driving is an active area of research in academia and industry. [ 22 ] Loon explored deep RL for autonomously navigating their high-altitude balloons. [ 23 ]

## Algorithms

Various techniques exist to train policies to solve tasks with deep reinforcement learning algorithms, each having their own benefits. At the highest level, there is a distinction between model-based and model-free reinforcement learning, which refers to whether the algorithm attempts to learn a forward model of the environment dynamics.

In model-based deep reinforcement learning algorithms, a forward model of the environment dynamics is estimated, usually by supervised learning using a neural network. Then, actions are obtained by using model predictive control using the learned model. Since the true environment dynamics will usually diverge from the learned dynamics, the agent re-plans often when carrying out actions in the environment. The actions selected may be optimized using Monte Carlo methods such as the cross-entropy method , or a combination of model-learning with model-free methods.

In model-free deep reinforcement learning algorithms, a policy  $\pi(a|s)$  is learned without explicitly modeling the forward dynamics. A policy can be optimized to maximize returns by directly estimating the policy gradient [ 24 ] but suffers from high variance, making it impractical for use with function approximation in deep RL. Subsequent algorithms have been developed for more stable learning and widely applied. [ 25 ] [ 26 ] Another class of model-free deep reinforcement learning algorithms rely on dynamic programming , inspired by temporal difference learning and Q-learning . In discrete action spaces, these algorithms usually learn a neural network Q-function  $Q(s,a)$  that estimates the future returns taking action  $a$  from state  $s$  . [ 14 ] In continuous spaces, these algorithms often learn both a value estimate and a policy. [ 27 ] [ 28 ] [ 29 ]

## Research

Deep reinforcement learning is an active area of research, with several lines of inquiry.

### Exploration

An RL agent must balance the exploration/exploitation tradeoff: the problem of deciding whether to pursue actions that are already known to yield high rewards or explore other actions in order to discover higher rewards. RL agents usually collect data with some type of stochastic policy, such as a Boltzmann distribution in discrete action spaces or a Gaussian distribution in continuous action spaces, inducing basic exploration behavior. The idea behind novelty-based, or curiosity-driven, exploration is giving the agent a motive to explore unknown outcomes in order to find the best solutions. This is done by "modify[ing] the loss function (or even the network architecture) by adding terms to incentivize exploration". [ 30 ] An agent may also be aided in exploration by utilizing demonstrations of successful trajectories, or reward-shaping, giving an agent intermediate rewards that are customized to fit the task it is attempting to complete. [ 31 ]

### Off-policy reinforcement learning

An important distinction in RL is the difference between on-policy algorithms that require evaluating or improving the policy that collects data, and off-policy algorithms that can learn a policy from data generated by an arbitrary policy. Generally, value-function based methods such as Q-learning are better suited for off-policy learning and have better sample-efficiency - the amount of data required to learn a task is reduced because data is re-used for learning. At the extreme, offline (or "batch") RL considers learning a policy from a fixed dataset without additional interaction with the environment.

#### Inverse reinforcement learning

Inverse RL refers to inferring the reward function of an agent given the agent's behavior. Inverse reinforcement learning can be used for learning from demonstrations (or apprenticeship learning) by inferring the demonstrator's reward and then optimizing a policy to maximize returns with RL. Deep learning approaches have been used for various forms of imitation learning and inverse RL. [ 32 ]

#### Goal-conditioned reinforcement learning

Another active area of research is in learning goal-conditioned policies, also called contextual or universal policies  $\pi(a|s, g)$  that take in an additional goal  $g$  as input to communicate a desired aim to the agent. [ 33 ] Hindsight experience replay is a method for goal-conditioned RL that involves storing and learning from previous failed attempts to complete a task. [ 34 ] While a failed attempt may not have reached the intended goal, it can serve as a lesson for how achieve the unintended result through hindsight relabeling.

#### Multi-agent reinforcement learning

Many applications of reinforcement learning do not involve just a single agent, but rather a collection of agents that learn together and co-adapt. These agents may be competitive, as in many games, or cooperative as in many real-world multi-agent systems. Multi-agent reinforcement learning studies the problems introduced in this setting.

#### Generalization

The promise of using deep learning tools in reinforcement learning is generalization: the ability to operate correctly on previously unseen inputs. For instance, neural networks trained for image recognition can recognize that a picture contains a bird even it has never seen that particular image or even that particular bird. Since deep RL allows raw data (e.g. pixels) as input, there is a reduced need to predefine the environment, allowing the model to be generalized to multiple applications. With this layer of abstraction, deep reinforcement learning algorithms can be designed in a way that allows them to be general and the same model can be used for different tasks. [ 35 ] One method of increasing the ability of policies trained with deep RL policies to generalize is to incorporate representation learning .

#### References