

Title: Meta-learning (computer science)

URL: [https://en.wikipedia.org/wiki/Meta-learning\\_\(computer\\_science\)](https://en.wikipedia.org/wiki/Meta-learning_(computer_science))

PageID: 4615464

Categories: Category:Machine 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

Meta-learning [ 1 ] [ 2 ] is a subfield of machine learning where automatic learning algorithms are applied to metadata about machine learning experiments. As of 2017, the term had not found a standard interpretation, however the main goal is to use such metadata to understand how automatic learning can become flexible in solving learning problems, hence to improve the performance of existing learning algorithms or to learn (induce) the learning algorithm itself, hence the alternative term learning to learn . [ 1 ]

Flexibility is important because each learning algorithm is based on a set of assumptions about the data, its inductive bias . [ 3 ] This means that it will only learn well if the bias matches the learning problem. A learning algorithm may perform very well in one domain, but not on the next. This poses strong restrictions on the use of machine learning or data mining techniques, since the relationship

between the learning problem (often some kind of database ) and the effectiveness of different learning algorithms is not yet understood.

By using different kinds of metadata, like properties of the learning problem, algorithm properties (like performance measures), or patterns previously derived from the data, it is possible to learn, select, alter or combine different learning algorithms to effectively solve a given learning problem. Critiques of meta-learning approaches bear a strong resemblance to the critique of metaheuristic , a possibly related problem. A good analogy to meta-learning, and the inspiration for Jürgen Schmidhuber 's early work (1987) [ 1 ] and Yoshua Bengio et al.'s work (1991), [ 4 ] considers that genetic evolution learns the learning procedure encoded in genes and executed in each individual's brain. In an open-ended hierarchical meta-learning system [ 1 ] using genetic programming , better evolutionary methods can be learned by meta evolution, which itself can be improved by meta meta evolution, etc. [ 1 ]

#### Definition

A proposed definition [ 5 ] for a meta-learning system combines three requirements:

The system must include a learning subsystem.

Experience is gained by exploiting meta knowledge extracted in a previous learning episode on a single dataset, or from different domains.

in a previous learning episode on a single dataset, or

from different domains.

Learning bias must be chosen dynamically.

Bias refers to the assumptions that influence the choice of explanatory hypotheses [ 6 ] and not the notion of bias represented in the bias-variance dilemma . Meta-learning is concerned with two aspects of learning bias.

Declarative bias specifies the representation of the space of hypotheses, and affects the size of the search space (e.g., represent hypotheses using linear functions only).

Procedural bias imposes constraints on the ordering of the inductive hypotheses (e.g., preferring smaller hypotheses). [ 7 ]

#### Common approaches

There are three common approaches: [ 8 ]

using (cyclic) networks with external or internal memory (model-based)

learning effective distance metrics (metrics-based)

explicitly optimizing model parameters for fast learning (optimization-based).

#### Model-Based

Model-based meta-learning models updates its parameters rapidly with a few training steps, which can be achieved by its internal architecture or controlled by another meta-learner model. [ 8 ]

#### Memory-Augmented Neural Networks

A Memory-Augmented Neural Network , or MANN for short, is claimed to be able to encode new information quickly and thus to adapt to new tasks after only a few examples. [ 9 ]

#### Meta Networks

Meta Networks (MetaNet) learns a meta-level knowledge across tasks and shifts its inductive biases via fast parameterization for rapid generalization. [ 10 ]

#### Metric-Based

The core idea in metric-based meta-learning is similar to nearest neighbors algorithms, which weight is generated by a kernel function. It aims to learn a metric or distance function over objects. The notion of a good metric is problem-dependent. It should represent the relationship between

inputs in the task space and facilitate problem solving. [ 8 ]

#### Convolutional Siamese Neural Network

Siamese neural network is composed of two twin networks whose output is jointly trained. There is a function above to learn the relationship between input data sample pairs. The two networks are the same, sharing the same weight and network parameters. [ 11 ]

#### Matching Networks

Matching Networks learn a network that maps a small labelled support set and an unlabelled example to its label, obviating the need for fine-tuning to adapt to new class types. [ 12 ]

#### Relation Network

The Relation Network (RN), is trained end-to-end from scratch. During meta-learning, it learns to learn a deep distance metric to compare a small number of images within episodes, each of which is designed to simulate the few-shot setting. [ 13 ]

#### Prototypical Networks

Prototypical Networks learn a metric space in which classification can be performed by computing distances to prototype representations of each class. Compared to recent approaches for few-shot learning, they reflect a simpler inductive bias that is beneficial in this limited-data regime, and achieve satisfied results. [ 14 ]

#### Optimization-Based

What optimization-based meta-learning algorithms intend for is to adjust the optimization algorithm so that the model can be good at learning with a few examples. [ 8 ]

#### LSTM Meta-Learner

LSTM -based meta-learner is to learn the exact optimization algorithm used to train another learner neural network classifier in the few-shot regime. The parametrization allows it to learn appropriate parameter updates specifically for the scenario where a set amount of updates will be made, while also learning a general initialization of the learner (classifier) network that allows for quick convergence of training. [ 15 ]

#### Temporal Discreteness

Model-Agnostic Meta-Learning (MAML) is a fairly general optimization algorithm , compatible with any model that learns through gradient descent. [ 16 ]

#### Reptile

Reptile is a remarkably simple meta-learning optimization algorithm, given that both of its components rely on meta-optimization through gradient descent and both are model-agnostic. [ 17 ]

#### Examples

Some approaches which have been viewed as instances of meta-learning:

Recurrent neural networks (RNNs) are universal computers. In 1993, Jürgen Schmidhuber showed how "self-referential" RNNs can in principle learn by backpropagation to run their own weight change algorithm, which may be quite different from backpropagation. [ 18 ] In 2001, Sepp Hochreiter & A.S. Younger & P.R. Conwell built a successful supervised meta-learner based on Long short-term memory RNNs. It learned through backpropagation a learning algorithm for quadratic functions that is much faster than backpropagation. [ 19 ] [ 2 ] Researchers at Deepmind (Marcin Andrychowicz et al.) extended this approach to optimization in 2017. [ 20 ]

In the 1990s, Meta Reinforcement Learning or Meta RL was achieved in Schmidhuber's research group through self-modifying policies written in a universal programming language that contains special instructions for changing the policy itself. There is a single lifelong trial. The goal of the RL agent is to maximize reward. It learns to accelerate reward intake by continually improving its own learning algorithm which is part of the "self-referential" policy. [ 21 ] [ 22 ]

An extreme type of Meta Reinforcement Learning is embodied by the Gödel machine , a theoretical construct which can inspect and modify any part of its own software which also contains a general theorem prover . It can achieve recursive self-improvement in a provably optimal way. [ 23 ] [ 2 ]

Model-Agnostic Meta-Learning (MAML) was introduced in 2017 by Chelsea Finn et al. [ 16 ] Given a sequence of tasks, the parameters of a given model are trained such that few iterations of gradient descent with few training data from a new task will lead to good generalization performance on that task. MAML "trains the model to be easy to fine-tune." [ 16 ] MAML was successfully applied to few-shot image classification benchmarks and to policy-gradient-based reinforcement learning. [ 16 ]

Variational Bayes-Adaptive Deep RL (VariBAD) was introduced in 2019. [ 24 ] While MAML is optimization-based, VariBAD is a model-based method for meta reinforcement learning, and leverages a variational autoencoder to capture the task information in an internal memory, thus conditioning its decision making on the task.

When addressing a set of tasks, most meta learning approaches optimize the average score across all tasks. Hence, certain tasks may be sacrificed in favor of the average score, which is often unacceptable in real-world applications. By contrast, Robust Meta Reinforcement Learning (RoML) focuses on improving low-score tasks, increasing robustness to the selection of task. [ 25 ] RoML works as a meta-algorithm, as it can be applied on top of other meta learning algorithms (such as MAML and VariBAD) to increase their robustness. It is applicable to both supervised meta learning and meta reinforcement learning .

Discovering meta-knowledge works by inducing knowledge (e.g. rules) that expresses how each learning method will perform on different learning problems. The metadata is formed by characteristics of the data (general, statistical, information-theoretic,... ) in the learning problem, and characteristics of the learning algorithm (type, parameter settings, performance measures,...). Another learning algorithm then learns how the data characteristics relate to the algorithm characteristics. Given a new learning problem, the data characteristics are measured, and the performance of different learning algorithms are predicted. Hence, one can predict the algorithms best suited for the new problem.

Stacked generalisation works by combining multiple (different) learning algorithms. The metadata is formed by the predictions of those different algorithms. Another learning algorithm learns from this metadata to predict which combinations of algorithms give generally good results. Given a new learning problem, the predictions of the selected set of algorithms are combined (e.g. by (weighted) voting) to provide the final prediction. Since each algorithm is deemed to work on a subset of problems, a combination is hoped to be more flexible and able to make good predictions.

Boosting is related to stacked generalisation, but uses the same algorithm multiple times, where the examples in the training data get different weights over each run. This yields different predictions, each focused on rightly predicting a subset of the data, and combining those predictions leads to better (but more expensive) results.

Dynamic bias selection works by altering the inductive bias of a learning algorithm to match the given problem. This is done by altering key aspects of the learning algorithm, such as the hypothesis representation, heuristic formulae, or parameters. Many different approaches exist.

Inductive transfer studies how the learning process can be improved over time. Metadata consists of knowledge about previous learning episodes and is used to efficiently develop an effective hypothesis for a new task. A related approach is called learning to learn , in which the goal is to use acquired knowledge from one domain to help learning in other domains.

Other approaches using metadata to improve automatic learning are learning classifier systems , case-based reasoning and constraint satisfaction .

Some initial, theoretical work has been initiated to use Applied Behavioral Analysis as a foundation for agent-mediated meta-learning about the performances of human learners, and adjust the instructional course of an artificial agent. [ 26 ]

AutoML such as Google Brain's "AI building AI" project, which according to Google briefly exceeded existing ImageNet benchmarks in 2017. [ 27 ] [ 28 ]

#### References

#### External links

Metalearning article in Scholarpedia

Vilalta, R.; Drissi, Y. (2002). "A perspective view and survey of meta-learning" (PDF) . Artificial Intelligence Review . 18 (2): 77– 95. doi : 10.1023/A:1019956318069 .

Giraud-Carrier, C.; Keller, J. (2002). "Meta-Learning". In Meij, J. (ed.). Dealing with the data flood . The Hague: STT/Beweton.

Brazdil, P.; Giraud-Carrier, C.; Soares, C.; Vilalta, R. (2009). "Metalearning: Concepts and Systems". Metalearning: applications to data mining . Springer. ISBN 978-3-540-73262-4 .

Video courses about Meta-Learning with step-by-step explanation of MAML , Prototypical Networks , and Relation Networks .