Title: Multi-agent reinforcement learning

URL: https://en.wikipedia.org/wiki/Multi-agent_reinforcement_learning

PageID: 62285602

Categories: Category:Deep learning, Category:Multi-agent systems, Category:Reinforcement learning

---

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

Multi-agent reinforcement learning (MARL) is a sub-field of reinforcement learning . It focuses on studying the behavior of multiple learning agents that coexist in a shared environment. [1] Each agent is motivated by its own rewards, and does actions to advance its own interests; in some environments these interests are opposed to the interests of other agents, resulting in complex group dynamics .

Multi-agent reinforcement learning is closely related to game theory and especially repeated games , as well as multi-agent systems . Its study combines the pursuit of finding ideal algorithms that maximize rewards with a more sociological set of concepts. While research in single-agent reinforcement learning is concerned with finding the algorithm that gets the biggest number of points for one agent, research in multi-agent reinforcement learning evaluates and quantifies social

metrics, such as cooperation, [2] reciprocity, [3] equity, [4] social influence, [5] language [6] and discrimination. [7]

Definition

Similarly to single-agent reinforcement learning , multi-agent reinforcement learning is modeled as some form of a Markov decision process (MDP) . Fix a set of agents I = { 1 , . . . , N } {\displaystyle I=\{1,...,N\}} . We then define:

A set S {\displaystyle S} of environment states.

One set A i {\displaystyle {\mathcal {A}}_{i}} of actions for each of the agents i ∈ I = { 1 , … , N } {\displaystyle i\in I=\{1,\dots ,N\}} .

P a → (s , s ′) = Pr (s t + 1 = s ′ ■ s t = s , a → t = a →) {\displaystyle P_{\vec {a}}(s,s')=\Pr(s_{t+1}=s'\mid s_{t}=s,{\vec {a}}_{t}={\vec {a}})} is the probability of transition (at time t {\displaystyle t}) from state s {\displaystyle s} to state s ′ {\displaystyle s'} under joint action a → {\displaystyle {\vec {a}}} .

R → a → (s , s ′) {\displaystyle {\vec {R}}_{\vec {a}}(s,s')} is the immediate joint reward after the transition from s {\displaystyle s} to s ′ {\displaystyle s'} with joint action a → {\displaystyle {\vec {a}}} .

In settings with perfect information , such as the games of chess and Go , the MDP would be fully observable. In settings with imperfect information, especially in real-world applications like self-driving cars , each agent would access an observation that only has part of the information about the current state. In the partially observable setting, the core model is the partially observable stochastic game in the general case, and the decentralized POMDP in the cooperative case.

Cooperation vs. competition

When multiple agents are acting in a shared environment their interests might be aligned or misaligned. MARL allows exploring all the different alignments and how they affect the agents' behavior:

In pure competition settings , the agents' rewards are exactly opposite to each other, and therefore they are playing against each other.

Pure cooperation settings are the other extreme, in which agents get the exact same rewards, and therefore they are playing with each other.

Mixed-sum settings cover all the games that combine elements of both cooperation and competition.

Pure competition settings

When two agents are playing a zero-sum game , they are in pure competition with each other. Many traditional games such as chess and Go fall under this category, as do two-player variants of video games like StarCraft . Because each agent can only win at the expense of the other agent, many complexities are stripped away. There is no prospect of communication or social dilemmas, as neither agent is incentivized to take actions that benefit its opponent.

The Deep Blue [8] and AlphaGo projects demonstrate how to optimize the performance of agents in pure competition settings.

One complexity that is not stripped away in pure competition settings is autocurricula . As the agents' policy is improved using self-play , multiple layers of learning may occur.

Pure cooperation settings

MARL is used to explore how separate agents with identical interests can communicate and work together. Pure cooperation settings are explored in recreational cooperative games such as Overcooked , [9] as well as real-world scenarios in robotics . [10]

In pure cooperation settings all the agents get identical rewards, which means that social dilemmas do not occur.

In pure cooperation settings, oftentimes there are an arbitrary number of coordination strategies, and agents converge to specific "conventions" when coordinating with each other. The notion of conventions has been studied in language [11] and also alluded to in more general multi-agent collaborative tasks. [12] [13] [14] [15]

Mixed-sum settings

Most real-world scenarios involving multiple agents have elements of both cooperation and competition. For example, when multiple self-driving cars are planning their respective paths, each of them has interests that are diverging but not exclusive: Each car is minimizing the amount of time it's taking to reach its destination, but all cars have the shared interest of avoiding a traffic collision . [17]

Zero-sum settings with three or more agents often exhibit similar properties to mixed-sum settings, since each pair of agents might have a non-zero utility sum between them.

Mixed-sum settings can be explored using classic matrix games such as prisoner's dilemma , more complex sequential social dilemmas , and recreational games such as Among Us , [18] Diplomacy [19] and StarCraft II . [20] [21]

Mixed-sum settings can give rise to communication and social dilemmas.

Social dilemmas

As in game theory , much of the research in MARL revolves around social dilemmas , such as prisoner's dilemma , [22] chicken and stag hunt . [23]

While game theory research might focus on Nash equilibria and what an ideal policy for an agent would be, MARL research focuses on how the agents would learn these ideal policies using a trial-and-error process. The reinforcement learning algorithms that are used to train the agents are maximizing the agent's own reward; the conflict between the needs of the agents and the needs of the group is a subject of active research. [24]

Various techniques have been explored in order to induce cooperation in agents: Modifying the environment rules, [25] adding intrinsic rewards, [4] and more.

Sequential social dilemmas

Social dilemmas like prisoner's dilemma, chicken and stag hunt are "matrix games". Each agent takes only one action from a choice of two possible actions, and a simple 2x2 matrix is used to describe the reward that each agent will get, given the actions that each agent took.

In humans and other living creatures, social dilemmas tend to be more complex. Agents take multiple actions over time, and the distinction between cooperating and defecting is not as clear cut as in matrix games. The concept of a sequential social dilemma (SSD) was introduced in 2017 [26] as an attempt to model that complexity. There is ongoing research into defining different kinds of SSDs and showing cooperative behavior in the agents that act in them. [27]

Autocurricula

An autocurriculum [28] (plural: autocurricula) is a reinforcement learning concept that's salient in multi-agent experiments. As agents improve their performance, they change their environment; this change in the environment affects themselves and the other agents. The feedback loop results in several distinct phases of learning, each depending on the previous one. The stacked layers of learning are called an autocurriculum. Autocurricula are especially apparent in adversarial settings, [29] where each group of agents is racing to counter the current strategy of the opposing group.

The Hide and Seek game is an accessible example of an autocurriculum occurring in an adversarial setting. In this experiment, a team of seekers is competing against a team of hiders. Whenever one of the teams learns a new strategy, the opposing team adapts its strategy to give the best possible counter. When the hiders learn to use boxes to build a shelter, the seekers respond by learning to use a ramp to break into that shelter. The hiders respond by locking the ramps, making them unavailable for the seekers to use. The seekers then respond by "box surfing", exploiting a glitch in the game to penetrate the shelter. Each "level" of learning is an emergent phenomenon, with the

previous level as its premise. This results in a stack of behaviors, each dependent on its predecessor.

Autocurricula in reinforcement learning experiments are compared to the stages of the evolution of life on Earth and the development of human culture . A major stage in evolution happened 2-3 billion years ago, when photosynthesizing life forms started to produce massive amounts of oxygen , changing the balance of gases in the atmosphere. [30] In the next stages of evolution, oxygen-breathing life forms evolved, eventually leading up to land mammals and human beings. These later stages could only happen after the photosynthesis stage made oxygen widely available. Similarly, human culture could not have gone through the Industrial Revolution in the 18th century without the resources and insights gained by the agricultural revolution at around 10,000 BC. [31]

Applications

Multi-agent reinforcement learning has been applied to a variety of use cases in science and industry:

Broadband cellular networks such as 5G [32]

Content caching [32]

Packet routing [32]

Computer vision [33]

Network security [32]

Transmit power control [32]

Computation offloading [32]

Language evolution research [34]

Global health [35]

Integrated circuit design [36]

Internet of Things [32]

Microgrid energy management [37]

Multi-camera control [38]

Autonomous vehicles [39]

Sports analytics [40]

Traffic control [41] (Ramp metering [42])

Unmanned aerial vehicles [43] [32]

Wildlife conservation [44]

AI alignment

Multi-agent reinforcement learning has been used in research into AI alignment . The relationship between the different agents in a MARL setting can be compared to the relationship between a human and an AI agent. Research efforts in the intersection of these two fields attempt to simulate possible conflicts between a human's intentions and an AI agent's actions, and then explore which variables could be changed to prevent these conflicts. [45] [46]

Limitations

There are some inherent difficulties about multi-agent deep reinforcement learning . [47] The environment is not stationary anymore, thus the Markov property is violated: transitions and rewards do not only depend on the current state of an agent.

Further reading

References

Yang, Yaodong; Wang, Jun (2020). "An Overview of Multi-Agent Reinforcement Learning from Game Theoretical Perspective". arXiv : 2011.00583 [cs.MA].

Kaiqing Zhang, Zhuoran Yang, Tamer Basar. Multi-agent reinforcement learning: A selective overview of theories and algorithms . Studies in Systems, Decision and Control, Handbook on RL and Control, 2021. [1]

Stefano V. Albrecht, Filippos Christianos, Lukas Schäfer. Multi-Agent Reinforcement Learning: Foundations and Modern Approaches . MIT Press, 2024. https://www.marl-book.com