+33 (0)1 48 10 09 46
formation@ambient-it.net
140 Avenue Jean Lolive 93500 Pantin - France

Updated 03/15/2024

Sign up

Introduction to Deep Reinforcement Learning

2 days (14 hours)

Presentation

Reinforcement Learning involves a large-scale system in which an agent must learn to solve a problem based on rewards. While this field has been around for some time, the arrival of Deep Learning has turned it on its head by making new tools available, approximating tools (Q function, policy, etc.) with neural networks. Numerous success stories have demonstrated that, despite its particular difficulty, this approach can revolutionize certain problems: video games, process optimization, the game of Go, continuous control or robotics.

The aim here is to present the basics of Reinforcement Learning, followed by the main advances that have appeared in recent years: Deep Q Learning, Rainbow, Policy gradients (A3C, PPO), exploration (World models, Imagination augmented agents) through to a detailed study of AlphaGo and AlphaGo Zero.

Objectives

- Mastery of reinforcement learning concepts and the main model-free approaches.
- Understanding exploration-based approaches and studying optimization approaches
- Model-based solutions: learning the model or using it directly
- Illustration of the points covered by the AlphaGo and AlphaGoZero application examples

Target audience

Developers, Architects, Big Data Data analyst / Data scientist & Engineer

Prerequisites

- Knowledge of Python

Further information

- As an introduction to Artificial Intelligence, we offer you the following training course
- Complementary technologies
  - Pytorch from Facebook
  - TensorFlow from Google

Deep Reinforcement Learning training program

[DAY 1]

1. Introduction to Reinforcement Learning concepts

- Introduction to reinforcement learning: controlling an agent in an environment defined by a state and possible actions. Fundamental modeling
- Markov Decision Processes modeling, Value Functions definition, Bellman equation, dynamic programming. Distinction between observation and state of the environment
- Value prediction approach: Temporal Difference & Monte Carlo. Example of these algorithms
- Policy iteration & evaluation: fundamental algorithm for policy convergence.
- Q Learning

2. Model Free Deep Reinforcement Learning (two examples of Tensorflow or PyTorch implementation are studied according to students' directions)

- Deep Q-Learning: Fundamental approach, Q-function approximation, Experience Replay, Double Q Learning. Detailed results study
- Deep Recurrent Q-Learning: The problem of a partially observable state. Comparison with Deep Q Learning
- Rainbow: analysis of advances and architectural modifications in Deep Q Learning: dueling networks, prioritized experience replay, distributional approach, use of noise. Analysis of the combined and individual contributions of each approach.

References: - Playing Atari with Deep Reinforcement Learning, Mnih et al, 2013. - Deep Recurrent Q-Learning for Partially Observable MDPs, Hausknecht and Stone, 2015 - Rainbow: Combining Improvements in Deep Reinforcement Learning, Hessel et al, 2017.

- Policy Gradients : Architecture Actor Critic

- Asynchronous A3C approach. Asynchronous definition of Deep Q Learning. A3C algorithm, interest, performance and flexibility of the asynchronous approach.
- Policy evolution by policy gradient: Trusted Policy Optimization and Proximal Policy Optimization. Advantages of the PPO approach. Study of results and application conditions.
- Soft actor critic: use of an entropy parameter to maximize exploration. Details architecture

References: - Asynchronous Methods for Deep Reinforcement Learning, Mnih et al, 2016 - Proximal Policy Optimization Algorithms, Schulman et al, 2017.

- Distributional approach: adaptation of equations and fundamental definitions. Motivation for the approach and observed results.
- Evolutionary algorithms: using Natural Evolution Strategies for Deep Convergence Reinforcement Learning. Vision of optimization and possible parallelization of learning. Analysis of comparative results.

References: - Soft Actor-Critic: Off-Policy Maximum Entropy Deep Reinforcement Learning with a Stochastic Actor, Haarnoja et al, 2018 - Evolution Strategies as a Scalable Alternative to Reinforcement Learning, Salimans et al, 2017

[DAY 2]

3. Exploring the environment

- Exploration versus learning: what weighting, what interest? How is exploration defined?
- Study of explorations based on a count of states/actions.
- Analysis of possible Hash state models. Hash learning with Variational Autoencoder (reminder of VAE principles)
- Curiosity" concepts
- Approach based solely on exploration without direct reward. Results, interests and discussions

References: - Exploration: A Study of Count-Based Exploration for Deep Reinforcement Learning, Tang et al, 2016 - Large-Scale Study of Curiosity-Driven Learning, Burda et al, 2018

4. Model based Deep Reinforcement Learning.

- Implementation of an agent-internal model to represent the environment.
- Study of different modeling strategies. Probabilistic or deterministic approach.
- Training a model in its "internal" environment and applying it to the target environment.
- Study of the "imagination" concept (Deepmind), Imagination Augmented Agent. Exploiting free learning with internal modeling of future states. Ablation studies.
- Comparative results

References: - Imagination-Augmented Agents for Deep Reinforcement Learning, Weber et al, 2017 - Recurrent World Models Facilitate Policy Evolution, Ha and Schmidhuber, 2018.

5. Model-based approaches: AlphaGo, AlphaGo Zero and derivatives

- Monte Carlo Tree Search (MCTS): analysis of the fundamental algorithm
- AlphaGo: four-stage learning analysis, and use of MCTS weighting the different neural networks available. Analysis of performance and results
- AlphaGo Zero: analysis of developments, use of MCTS in learning.
  AlphaGo VS AlphaGO Zero comparison
- AlphaZero: generalizing the AlphaGo Zero approach to other approaches
- Imitation Learning: definition and examples
- Expert Iteration: use of MCTS for internal modeling of an expert model to implement imitation learning.

References: - Mastering Chess and Shogi by Self-Play with a General Reinforcement Learning Algorithm, Silver et al, 2017 - Thinking Fast and Slow with Deep Learning and Tree Search, Anthony et al, 2017

6. Scaling RL learning and recent algorithms

- Analysis of GPU versus CPU parallelization possibilities. Approach and mitigation strategies. Data-efficiency vision of proposed approaches.
- Distributive approach for greater learning parallelization
- Analysis of the R2D2 algorithm: use of recurrent models and parallelization, in-depth analysis of biases induced by variation in the hidden state of the network.

References: - Accelerated Methods for Deep Reinforcement Learning, Stooke and Abbeel, 2018 - Recurrent Experience Replay in Distributed Reinforcement Learning, Kapturowski et al, 2018

Companies concerned

This course is aimed at both individuals and companies, large or small, wishing to train their teams in a new advanced computer technology, or to acquire specific business knowledge or modern methods.

Positioning on entry to training

Positioning at the start of training complies with Qualiopi quality criteria. As soon as registration is finalized, the learner receives a self-assessment questionnaire which enables us to assess his or her estimated level of proficiency in different types of technology, as well as his or her expectations and personal objectives for the training to come, within the limits imposed by the selected format. This questionnaire also enables us to anticipate any connection or security difficulties within the company (intra-company or virtual classroom) which could be problematic for the follow-up and smooth running of the training session.

Teaching methods

Practical course: 60% Practical, 40% Theory. Training material distributed in digital format to all participants.

Organization

The course alternates theoretical input from the trainer, supported by examples, with brainstorming sessions and group work.

Validation

At the end of the session, a multiple-choice questionnaire verifies the correct acquisition of skills.

Sanction

A certificate will be issued to each trainee who completes the course.