AlphaZero Algorithm - Implementations

03.19.2021

Vincent Manier

Main Take-aways from last week:

- There is no evidence that the network is learning
- IID assumptions -> Independent and Identically Distributed. IID could be broken by using the same sequences (Periodicite)
- Questions about the loss evolution?
  - Why does the loss spike?
  - Numerical instability
  - Stats sur la valeur des weights - max, moyenne?
  - Explosion des gradients? norme maximale
  - L2 regularization?
- Solutions:
  - Radians clipping
  - Batch normalization
- Tests:
  - Network on a stand-alone basis
  - Competition against Minimax
  - AlphaZero.jl has some redundancy measures to assess parties diversity
- Network monitoring:
  - Use pytorch to get a dictionary of weights / gradients
  - Use tensorboard
- Development:
  - It is important to test modules independently

A. Objectives

a. Change the Tic Tac Toe algorithm by:

   i. Storing experiences in memory after each self-play

   ii. Training the networks on a batch of diverse experiences every x iterations of self-play

   iii. Control if the network is learning based on a few easy board positions

   iv. Creating a competition module between old and new networks. If significantly better, replacing the old network with the new network for game generation.

b. Assess the tic-tac-toe player by competing against:

   i. A network-only

   ii. A MCTS-only

   iii. A mini-max

c. Modularize my implementation:

   i. Test MCTS on a stand-alone basis

B. Testing / Experiences

a. Use 2 board positions as a test case

```
game_state1 = np.array([[-1, 1, -1], [0, 1, 0], [0, 0, 0]])
```

-1	1	-1
0	1	0
0	0	0

```
game_state2 = np.array([[-1, 1, -1], [0, 1, -1], [0, 0, 0]])
```

-1	1	-1
0	1	-1
0	0	0

i. Fixing bug in my loss function

ii. Look at value / probability of each test position and see the evolution

b. Diversity of experiences was a major issue

i. Tends to repeat the same position!

ii. Symmetries

iii. Deduplication / averaging

iv. MCTS exploration parameters (temperature, puct)