Comparing Model-Free Deep Reinforcement Learning Algorithms on Stock Market

1. The Aim

To understand the performance difference of

model-free deep reinforcement learning algorithms
on **stock market**

in terms of

training speed
performance
generalizability

2. Methods

Comparing

Policy Optimization
Proximal Policy Optimization (PPO)

Q-Learning
(TD3) Twin Delayed DDPG

Observation
Balance
Current Portfolio
Stock Prices
Technical Indicators

Reward
Rate of Return

Actions
Buy/Sell Stocks

F : Ford Motor Company
NVDA : NVIDIA Corporation

Hyperparameter optimization
using Optuna Framework
on six AWS EC2 instances

Training frameworks
FinRL Library
Stable Baselines3

OpenAI Gym

3. Results

Algorithms (Stocks)	Cumulative Return	Annual Volatility	Sharpe Ratio
PPO (F, NVDA)	16.552928	0.468879	1.459257
TD3 (F, NVDA)	16.559846	0.468114	1.461074
PPO (GM, AMD)	31.933818	0.606478	1.451827
TD3 (GM, AMD)	39.035147	0.632949	1.476289

Q-Learning **generalizes better**.

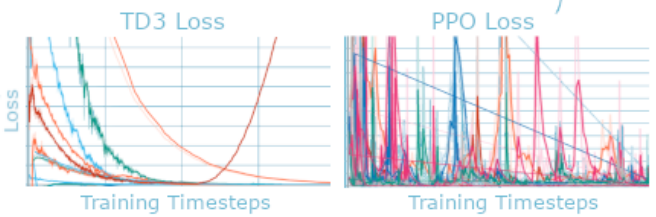
No statistical difference in performance → Overfitting?

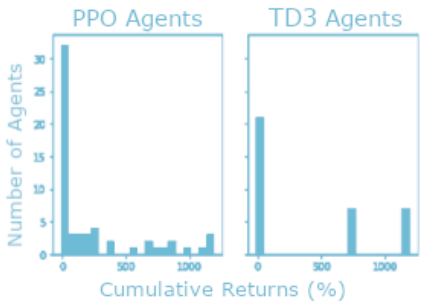
35 TD3 Agents and 59 PPO Agents in same time frame

TD3 is more **frame efficient**. PPO is **faster** in training.

TD3 is more **stable**.

PPO is **sensitive to hyperparameters**.





4. Contact

TU Delft

Murat Kaan Meral
M.K.Meral@student.tudelft.nl