Robustness Analysis of Deep RL for Portfolio Selection

Marc VELAY [1,2] Bich-Liên DOAN [1] Fabrice POPINEAU [1] Arpad RIMMEL [1] [1]LISN, [2]Lusis Payments
<firstname.lastname>@centralesupelec.fr

Sequential Decision Making under Uncertainty

- **Financial Markets** are very noisy, non-stationary environments
- **Portfolio Selection** aims to spread risk exposure across assets, over time to:
  - Statistically guarantee minimum returns
  - Avoid individual fluctuations for reduced variance

Online Portfolio Selection

- **Traditional approaches** – Rely on historical standard deviation, trends and asset correlation.
- **Limited impact** – Better than individual assets, but comparable in performance to following the market
- **Deep Reinforcement Learning** – Some promising results from recent works. Many claim to reach State of the Art performance, but few:
  - Compare to other DRL approaches
  - Analyze actions chosen
  - Analyze robustness to market conditions and drift

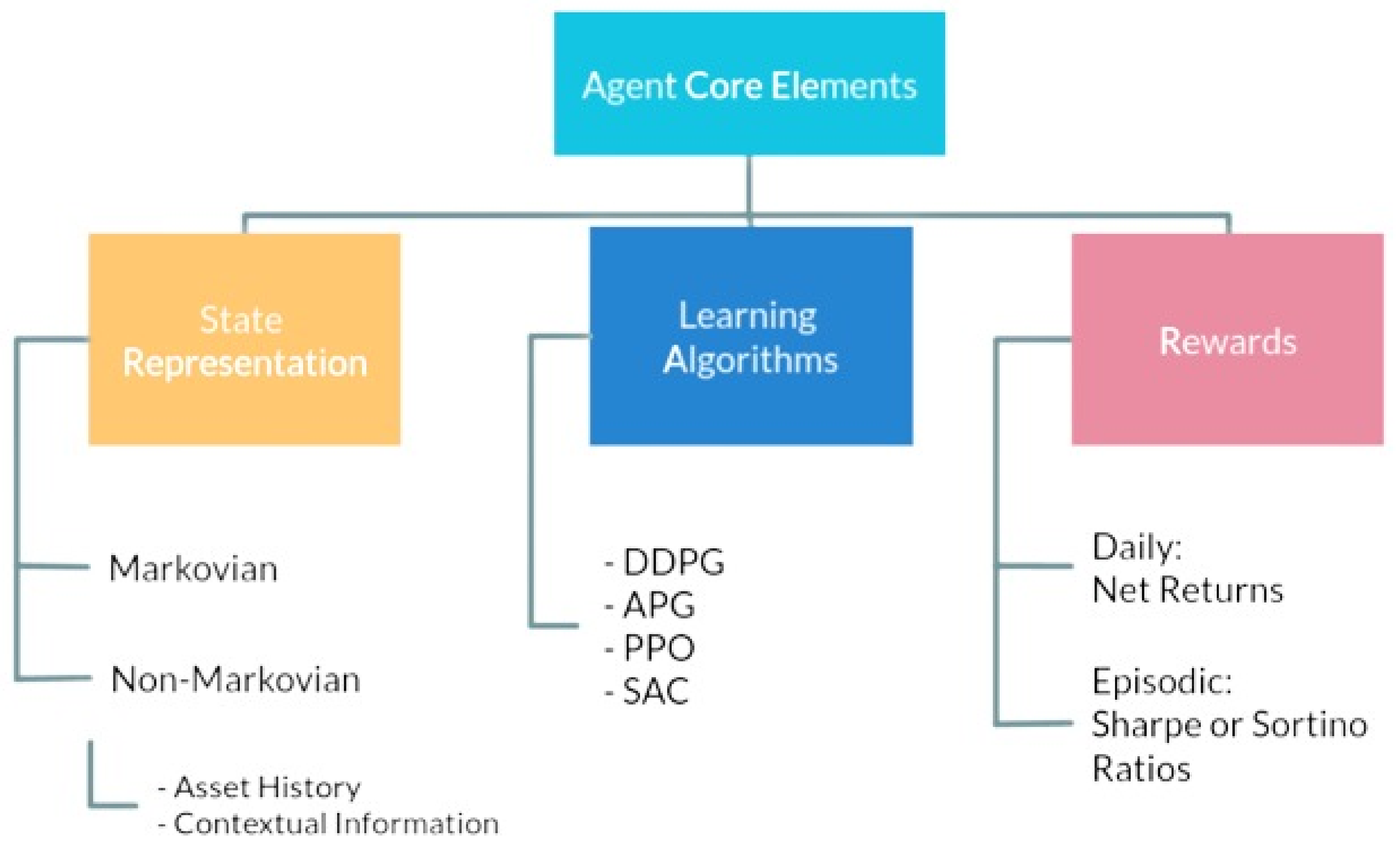


Figure 1. Diverging Implementation Choices

Experiments

Common setup The process relies on a combination of assets from the CAC40, commodities futures and ForEx pairs. The training and validation data is from 2002 to 2019, covering major historical events and various market conditions. Daily OHLCV (Open, High, Low, Close prices and Volume) bars are used as the base for any other representations used by agents. We use data from 2020 onwards for our backtesting strategies.

Metrics:

- **Performance** Net returns, the difference in value between the start and end of investments.
- **Risk** Sharpe or Sortino ratios [1], comparing portfolio returns to its fluctuations. We aim for more stable returns than one-off growths.
- **Robustness** Measures how the approach performs outside of nominal conditions [4]
  - CVaR, the expected returns in worse-case scenarios
  - Performance over-time, analyzing shorter time-frames at regular intervals, to observe variations in behavior w.r.t market conditions.

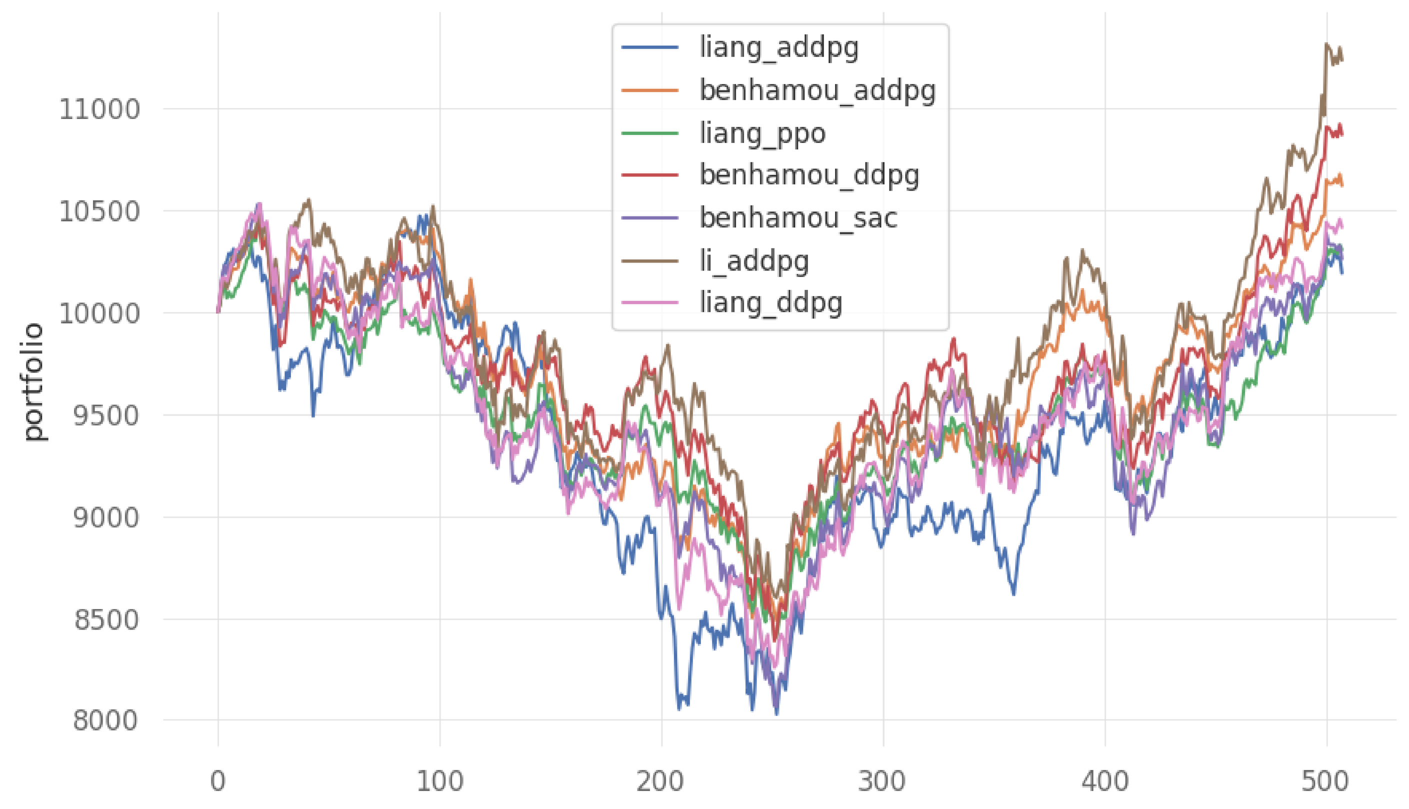
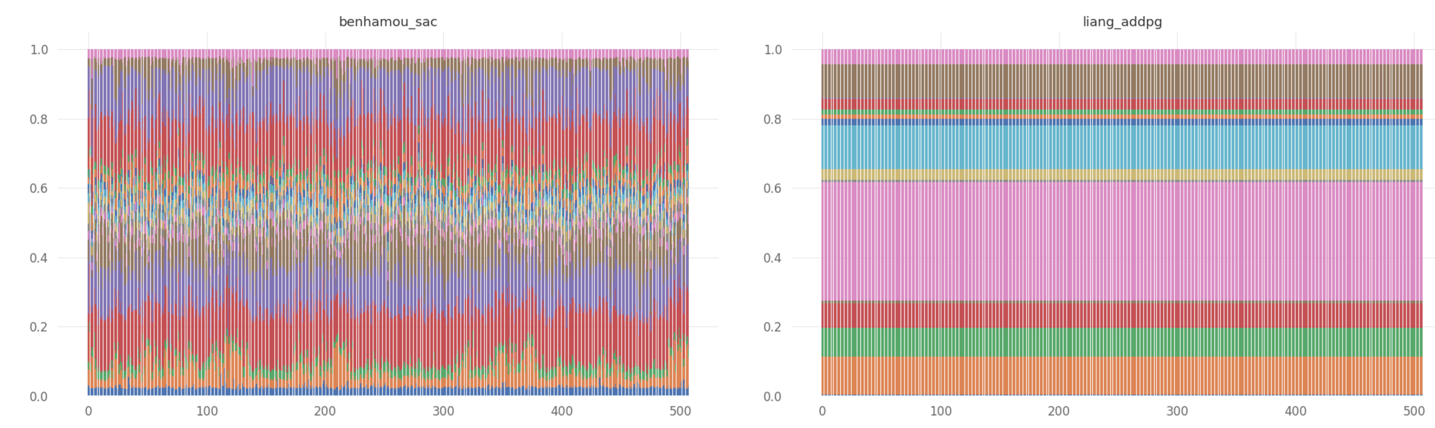


Figure 2. Sampled Approaches Performances

Results & Conclusions



(a) Frequent re-allocation (b) Constant allocation

Figure 3. Observed Agent Allocation Strategies

- **Problems encountered** large amounts of agent policies tend towards constant allocations, after hyperparameter selection. This is also observed in other works [2]. While agents do reach interesting performance results, this is not a long-term optimal behavior, which adapts to market conditions.
- **Solutions**
  - Change training data to be broader by sampling fewer assets from a much larger pool [5]
  - Use estimation models which work with small sample sizes
  - Employ transfer learning to fine-tune policies obtained through Inverse Reinforcement Learning
  - Use World Models [3] to fully train in simulations, only backtesting on real data

References

[1] Eric Benhamou et al. *Bridging the Gap Between Markowitz Planning and Deep Reinforcement Learning*. SSRN Scholarly Paper ID 3702112. Rochester, NY: Social Science Research Network, Sept. 30, 2020. DOI: 10.2139/ssrn.3702112.

[2] Ricard Durall. *Asset Allocation: From Markowitz to Deep Reinforcement Learning*. July 14, 2022. arXiv: 2208.07158[cs,q-fin].

[3] David Ha and Jürgen Schmidhuber. "World Models". In: (Mar. 28, 2018). DOI: 10.5281/zenodo.1207631. arXiv: 1803.10122[cs,stat].

[4] Janosch Moos et al. "Robust Reinforcement Learning: A Review of Foundations and Recent Advances". In: *Machine Learning and Knowledge Extraction* 4.1 (Mar. 2022), pp. 276–315. ISSN: 2504-4990. DOI: 10.3390/make4010013.

[5] Uta Pigorsch and Sebastian Schäfer. "High-Dimensional Stock Portfolio Trading with Deep Reinforcement Learning". In: *arXiv:2112.04755 [cs, q-fin]* (Dec. 9, 2021). arXiv: 2112.04755.