BRNO UNIVERSITY OF TECHNOLOGY
VYSOKÉ UČENÍ TECHNICKÉ V BRNĚ

FACULTY OF INFORMATION TECHNOLOGY
FAKULTA INFORMAČNÍCH TECHNOLOGIÍ

DEPARTMENT OF INTELLIGENT SYSTEMS
ÚSTAV INTELIGENTNÍCH SYSTÉMŮ

# REINFORCEMENT LEARNING FOR AUTOMATED STOCK PORTFOLIO ALLOCATION

VYUŽITÍ ZPĚTNOVAZEBNÉHO UČENÍ PRO AUTOMATICKOU ALOKACI AKCIOVÉHO PORTFOLIA

BACHELOR'S THESIS
BAKALÁŘSKÁ PRÁCE

AUTHOR ZDENĚK LAPEŠ
AUTOR PRÁCE

SUPERVISOR doc. RNDr. MILAN ČEŠKA, Ph.D.
VEDOUCÍ PRÁCE

BRNO 2023

BRNO FACULTY
UNIVERSITY OF INFORMATION
OF TECHNOLOGY TECHNOLOGY

# Bachelor's Thesis Assignment

148202

Institut: Department of Intelligent Systems (UITS)
Student: **Lapeš Zdeněk**
Programme: Information Technology
Specialization: Information Technology
Title: **Reinforcement Learning for Automated Stock Portfolio Allocation**
Category: Artificial Intelligence
Academic year: 2022/23

Assignment:

1. Study the state-of-the-art methods for automated stock portfolio allocation. Focus on the methods based on reinforcement learning and planning in Markov Decision Processes.
2. Experimentally evaluate selected open access tools for automated portfolio allocation including e.g. FinRL-Meta and identify their weak points.
3. Propose and implement improvements of a selected method/tool allowing to mitigate theseweak points.
4. Using suitable benchmarks and datasets, perform a detailed experimental evaluation of the implemented improvements with the focus on the portfolio allocation returns.

Literature:

Rao A., Jelvis T., Foundations of Reinforcement Learning with Applications in Finance. 1st Edition, Taylor & Francis 2022
* Li, Xinyi and Li, Yinchuan and Zhan, Yuancheng and Liu, Xiao-Yang, Optimistic Bull or Pessimistic Bear: Adaptive Deep Reinforcement Learning for Stock Portfolio Allocation, In ICML 2019.
* Liu X.-Y. Rui J. Gao J. aj.: FinRL-Meta: A Universe of Near-Real Market Environments for Data-Driven Deep Reinforcement Learning in Quantitative Finance. Workshop on Data Centric AI 35th Conference on Neural Information Processing Systems at NeurIPS 2021.
* Mao Guan and Xiao-Yang Liu. 2021. Explainable Deep Reinforcement Learning for Portfolio Management: An Empirical Approach. In ICAIF 2021.


Requirements for the semestral defence:
Items 1, 2, and partially 3.

Detailed formal requirements can be found at https://www.fit.vut.cz/study/theses/

Supervisor: **Češka Milan, doc. RNDr., Ph.D.**
Head of Department: Hanáček Petr, doc. Dr. Ing.
Beginning of work: 1.11.2022
Submission deadline: 10.5.2023
Approval date: 3.11.2022

## Abstract

Portfolio allocation is about selecting a set of assets to invest in. The goal is to maximize the expected return for a given level of risk for an investing horizon selected by us. In former times, this the process is usually done manually by an expert investor. Nowadays, there are many portfolio allocation methods that are not successful in the real world or can be improved and the potential of current technologies is not fully explored in the financial market. To solve the problem of portfolio allocation, I propose a methods based on reinforcement learning. The agent will learn the exact strategies of selecting assets and their weight for the portfolio based on which the human expert would select it, like fundamental analysis and the health of the company. The thesis deals with the problem of portfolio allocation using reinforcement learning, which helps in selecting the best assets and their importance for the portfolio.



## Abstrakt

Alokace portfolia se zaměřuje na výběr souboru aktiv, do kterých investujete. Cílem je maximalizovat očekávaný výnos pro danou míru rizika pro námi zvolený investiční horizont. V dřívějších dobách tohle proces je obvykle prováděn ručně zkušeným investorem. V dnešní době existuje mnoho metod alokace portfolia, které nejsou úspěšné v reálném světě nebo je lze zlepšit a potenciál současných technologií není na finančním trhu plně prozkoumán. Pro řešení problému alokace portfolia navrhuji metody založené na posilovacím učení. Agent se naučí přesné strategie výběru aktiv a jejich váhu pro portfolio, na základě kterého by je lidský expert vybíral, jako je fundamentální analýza a zdraví společnosti. Diplomová práce se zabývá problémem alokace portfolia pomocí posilovacího učení, které pomáhá při výběru nejlepších aktiv a jejich důležitosti pro portfolio.



## Keywords

artificial intelligence, AI, reinforcement learning, stock portfolio allocation, modern portfolio theory, Q-learning, neural networks, stock market

## Klíčová slova

umělá inteligence, AI, posilované učení, alokace akciového portfolia, moderni teorie portfolia, Q-learning, neuronové sítě, akciový trh


## Reference

LAPEŠ, Zdeněk. *Reinforcement Learning for Automated Stock Portfolio Allocation.* Brno, 2023. Bachelor’s thesis. Brno University of Technology, Faculty of Information Technology. Supervisor doc. RNDr. Milan Češka, Ph.D.

# Reinforcement Learning for Automated Stock Portfolio Allocation

## Declaration

I hereby declare that this Bachelor's thesis was prepared as an original work by the author under the supervision of Mr. Milan Češka, Ph.D. I have listed all the literary sources, publications and other sources, which were used during the preparation of this thesis.

. . . . . . . . . . . . . . . . . . . . . . .
Zdeněk Lapeš
April 4, 2023


## Acknowledgements

I would like to thank my supervisor, Mr. Milan Češka, for his guidance and support during the preparation of this thesis. I would also like to thank my family and friends for their support.

# Contents

# List of Figures

# Chapter 1

# Introduction

## 1.1 Background

The **Portfolio allocation problem** is to spread appropriate finite cash budget into financial instruments [2]. Under the financial instruments, we can imagine Stocks, Bonds, Mutual Funds, Commodities, Derivatives, Real Estate Investments Trusts (REITs), Exchange-Traded Funds (ETFs), and many more. The outcome should be to increase the initial capital over the course of a selected investing horizon, which can vary from a few days to decades. Portfolio management is essential for investors, particularly those who manage large sums of money such as institutional investors, pension funds, and wealthy individuals. While allocating assets instead of cash one must think about minimizing risk and maximizing the expected return on the investment. For that, the key considered strategy is **diversification**, which involves spreading investments across different instrument classes and markets in order to reduce the overall risk of the portfolio. A portfolio full of different assets can change over time due to market conditions, where the value of other assets may increase or decrease, which may cause the portfolio to become imbalanced. **Rebalancing** ensures that the portfolio remains aligned with the investor's goals and risk tolerance.

Among other portfolio allocation strategies could be mentioned **Modern portfolio theory (MPT)** to optimally allocate assets in a portfolio [6]. MPT uses statistical tools to determine the efficient frontier, which is the set of optimal portfolios that offer the highest expected return for a given level of risk, or the lowest risk for a given level of expected return. [7] Another approach is **Mean-variance optimization**, which uses mathematical models to determine the optimal portfolio based on an investor's risk tolerance and expected returns [4].

These approaches are not too appropriate for portfolio management, because the stock market is stochastic, volatile, quickly changing, and uncertain environment. These strategies are not flexible enough to adapt to the changing environment like the stock market, because they assume the future will be similar to the past, which may not always be accurate.

So, the most recent state-of-the-art portfolio management strategies are based on machine learning techniques. **Reinforcement learning (RL)** is a type of machine learning that is well-suited for solving problems involving decision-making and control [5]. In the context of portfolio allocation, RL can be used to optimize the allocation of assets in a portfolio in order to maximize returns or minimize risk. RL algorithms can learn from historical data and adapt to changing market conditions, which can lead to more efficient and profitable portfolio management. The benefits of RL have been used in many different fields, such as robotics, games, and finances.

In the last decade, RL has become popular, because of its ability to learn difficult tasks in a variety of domains without knowing the environment model [5]. RL has advantages, such as flexibility, adaptability, and utilization of various information like e.q. experience gained from the environment under certain conditions. The agent is trained under a certain policy in a particular environment, which is modeled using **Markov Decision Process (MDP)**. MDP is a mathematical framework for modeling sequential decision-making problems [3]. MDP can be used to model the fully observable environment, where the agent can observe the state of the environment. If the environment is not fully observable, then the agent can observe only a part of the state of the environment, which is called **partially observable Markov decision process (POMDP)** [5]. In finances, the environment is usually fully observable, because the agent can observe the state of the environment. MDP is composed of the following elements:

1. **State:** The state is the current situation of the environment.

2. **Action:** The action is the decision that the agent can take.

3. **Reward:** The reward is the feedback that the agent receives after taking an action.

4. **Transition:** The transition is the change of the state after taking an action.

In agent training we handle the following problems:

- **State space**
  The state space is a finite set of all possible configurations of the environment. In the context of portfolio allocation, the state space can be defined as the finite set of all possible instrument features (fundamental and technical analysis) and their weights in the portfolio.

- **Action space**
  Action space should be designed so that the agent weights the assets in the portfolio. Here the question is: Should be this asset in the portfolio and if yes, what is the weight of this asset in the portfolio? These decisions are crucial for the performance of the agent. It is really difficult to find the optimal policy for the portfolio allocation because the agent has to choose between multiple assets with various differences in information about the assets. Also, actions should be considered profitable and safe in the long term, which means that the agent usually has to make decisions based on long-term rewards or on the defined investment horizon.

- **Reward function**
  The reward should reflect the agent's performance in the environment. Is the current portfolio value increasing or decreasing after the agent takes actions proposed by the policy?

When the state space is too large, then is merely impossible to be explored with the limited computational resources **Deep Reinforcement Learning (DRL)** can be used. DRL is a subfield of Reinforcement Learning (RL) that combines the use of deep neural networks with RL algorithms. In traditional RL, the agent's policy and value functions are typically represented by simple, hand-designed features or a small number of parameters. In contrast, DRL uses deep neural networks to represent these functions, allowing the agent to learn from high-dimensional and complex inputs. DRL algorithms are used to train agents to perform a wide range of tasks, such as playing video games, controlling robotic arms, and driving cars. There are several popular algorithms in DRL, such as: **Deep Q-Network (DQN)**, **Deep Deterministic Policy Gradient (DDPG)**, **Proximal Policy Optimization (PPO)**, **Soft Actor-Critic (SAC)**, and **Twin Delayed Deep Deterministic Policy Gradient (TD3)**.

## 1.2 Limitations

1. **Data availability:** DRL models require large amounts of historical data to train effectively, which may be difficult to obtain for certain assets or markets.

2. **Model Overfitting:** DRL models can easily overfit to the training data, leading to poor performance on unseen data.

3. **High computational cost:** DRL models can require significant computational resources to train agents.

4. **Risk management:** DRL models may not be able to effectively handle risk management, such as different market situations (Market sentiment, Bull and Bear markets).

## 1.3 Aim of the Thesis

We will evaluate the performance of portfolio allocation methods based on DRL and compare them to traditional portfolio optimization techniques (MPT, Mean-Variance). Our goal is to determine the potential of DRL for portfolio allocation and identify the limitations of DRL-based portfolio allocation methods for future research.

The thesis objectives are:

- **Experimental evaluation & Benchmarks**
  Compare existing portfolio allocation agents. Evaluate the performance of the RL agents by comparing them with the baseline portfolio management strategies, such as MPT, Mean-Variance Optimization, and indexes (DJI, Nasdaq-100).

- **Dataset**
  Create a suitable dataset for the portfolio allocation problem. Datasets will be focused on the company's financial data, such as fundamental and technical analysis data.

- **Reimplementation**
  Try to improve current agents (Portfolio Allocation agent from **FinRL** [1]) with new datasets, focusing on Data Engineering and different DRL algorithms.

The thesis will be implemented using the programming language Python3 and open-source libraries such as NumPy, Pandas, Stable Baselines3, and OpenAI Gym.

pen

# Chapter 2

# Preliminaries

TODO

# Chapter 3

# Related Work

TODO

# Chapter 4

# Methodology

TODO

# Chapter 5

# Results

TODO

# Chapter 6

# Discussion

TODO

# Chapter 7

# Conclusion

**TODO**

# Chapter 8

# Future Work

TODO

# Chapter 9

# Acknowledgements

TODO

# Chapter 10

# References

TODO

# Chapter 11

# Appendix

TODO

# Bibliography


[1] Guan, M. and Liu, X.-Y. *Explainable Deep Reinforcement Learning for Portfolio Management: An Empirical Approach.* arXiv, 2021. DOI: 10.48550/ARXIV.2111.03995. Available at: https://arxiv.org/abs/2111.03995.

[2] Oshingbesan, A., Ajiboye, E., Kamashazi, P. and Mbaka, T. *Model-Free Reinforcement Learning for Asset Allocation.* arXiv, 2022. DOI: 10.48550/ARXIV.2209.10458. Available at: https://arxiv.org/abs/2209.10458.

[3] Rao, A. and Jelvis, T. *Foundations of Reinforcement Learning with Applications in Finance.* CRC Press, 2022. ISBN 9781000801101. Available at: https://books.google.cz/books?id=n_-VEAAAQBAJ.

[4] Soeryana, E., Fadhlina, N., Firman, S., Rusyaman, E. and Supian, S. Mean-variance portfolio optimization by using time series approaches based on logarithmic utility function. *IOP Conference Series: Materials Science and Engineering.* january 2017, vol. 166, p. 012003. DOI: 10.1088/1757-899X/166/1/012003.

[5] Sutton, R. and Barto, A. *Reinforcement Learning, second edition: An Introduction.* MIT Press, 2018. Adaptive Computation and Machine Learning series. ISBN 9780262039246. Available at: https://books.google.cz/books?id=5s-MEAAAQBAJ.

[6] Wikipedia contributors. *Modern portfolio theory — Wikipedia, The Free Encyclopedia.* 2022. [Online; accessed 28-December-2022]. Available at: https://en.wikipedia.org/w/index.php?title=Modern_portfolio_theory&oldid=1126978098.

[7] Širůček, M. and Křen, L. Application of Markowitz Portfolio Theory by Building Optimal Portfolio on the US Stock Market. *Acta Universitatis Agriculturae et Silviculturae Mendelianae Brunensis.* Mendel University Press. september 2015, vol. 63, no. 4, p. 1375–1386. DOI: 10.11118/actaun201563041375. Available at: http://dx.doi.org/10.11118/actaun201563041375.