澳門大學
UNIVERSIDADE DE MACAU
UNIVERSITY OF MACAU

科技學院
Faculdade de Ciências e Tecnologia
Faculty of Science and Technology

Enhancing Stock Trading Strategies with Integrated Feature Enhancement & Reward Shaping :

A Reinforcement Learning Approach

Student: Wong Kai Yuan (DC026157)

Supervisor: Prof Ryan Leong Hou U

Table of contents

01
Introduction

Motivation

Stock Trading

Often heard, but never deeply understood.

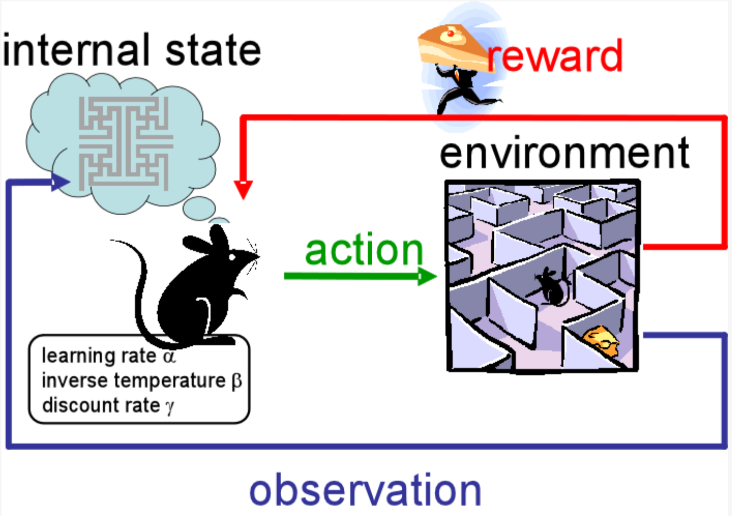
Advancement of AI

Recent years of rapid advancement of AI

Take such an opportunity to delve deeper and learn more about these topics as my FYP

Background of study

Reinforcement Learning



- Branch of Machine Learning
- Agent learns to make decision (will receive reward / penalty) by interacting with environment.
- Aiming to maximize cumulative rewards. Aka make best decision in long term.
- MDP Framework as mathematical framework: (States, Action, Rewards, Transition)

RL Application

- In various domains, most notable in mastering complex robotic problems and games at superhuman level, such as **AlphaGo**.
- RL agent able to develop a good strategy from complex and challenging environment.
- **This make me wonder if RL can perform well in the “game” of stock trading, or at least a good candidate**.

Problem Statement

Problem Statement

Limited Data

Need to admit that Stocks Trading Env is not so RL friendly. Unlike games with extensive rich data, stock trading data is limited to a few lines of prices per stock.

Features

Stock market data is characterized by:

- **Sparse:** Data are low significance.
- **Noisy**: Data has random fluctuations.
- **Nonstationary**: Statistical property of data is not consistent

Reward

- **Delayed reward**: The financial outcomes of trading decisions are often realized after a significant delay, complicating the association between actions and their consequences.
- **Sparse reward**: Profitable trades are infrequent, resulting in sparse rewards that make learning effective trading strategies difficult.

Study Focus

To address these challenges, it is crucial to improve the RL agent's ability to **perceive data** and the ability to **make decisions**. Therefore, our study focus on:

Feature Enhancement

Improving the representation of input data, we aim to provide the RL agent with more informative and relevant features.

Reward Shaping

By refining the reward function, we aim to provide more meaningful feedback to the RL agent. This encourages good behaviour and stops bad behaviour.

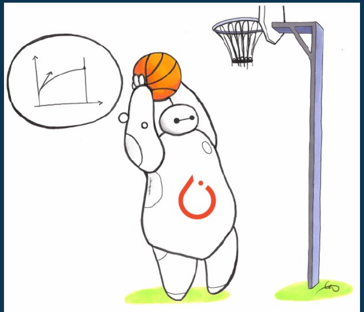
Test the Methodology Developed

Further test the same methodology on other stocks from varies sectors.

Tools Details

- **Environment:**

AnyTrading Gym with small customization



- **Algorithm:**

PPO from SB3

- **Data Chosen:**

Goldman Sachs Group Inc 10 Years daily data

(95% for training, 5% for testing)

02
Methodology of Feature Enhancement

Overview of Feature Enhancement

Methodology

- Original : Basic Data
- #1 : Extended Market Data
- #2 : Integrating Technical Indicators
- #3 : Integrating Candlestick Pattern

Original Feature Function

Only 2 Features for agent

- Close Price
- Price Difference

Testing Result

Performance Metric	Results
Total Profit :	0.80 (means loss 20%, 1.0 being breakeven)
Sharpe Ratio :	-1.61 (perform poorly in terms of risk-adjusted return)
Maximum Drawdown :	-0.35 (max drawdown was 35%)

Original Feature Function

Problem

- 2 Features as input are insufficient to capture the complex movements of the stock market.
- This simplicity led to poor performance metrics, indicating underfitting.

Proposed Solution

- Clear that the input features needed to be enriched.
- #1 Solution is adding all available **columns (like volume)** from the dataframe as the input feature.
- Our hypothesis is, this will make the model better at capturing more complex market patterns

#1 Feature Enhancement

Extended Market Data

- Dataset column include ["Open", "Close", "Low", "High", "Volume", "Adj Close"]
- Along with the day-over-day differences for each column
- Resulting in a total of 12 features.

Rationale

- Provide richer features for agent like Volume and Price Difference
- Model can potentially identify more complex patterns
- Hence, resulting in better predictive accuracy.

#1 Feature Enhancement

Testing Result

Performance Metric	Results
Total Profit :	1.19
Sharpe Ratio :	0.43
Maximum Drawdown :	-0.32

noticeable improvement !

Discussion

- Result supported the hypothesis where adding more data helps agent learn better.
- This motivate us to add more relevant features, which lead to #2 Feature Enhancement (add Technical Indicators)

#2 Feature Enhancement

Integrating Technical Indicators

- Technical Indicators have proven to be useful in financial analysis.

- Provide insights of market trend, momentum, and volatility **not direct** from raw price data.

#2 Feature Enhancement

Testing Result

Performance Metric	Results
Total Profit :	1.16
Sharpe Ratio :	0.52
Maximum Drawdown :	-0.32

Problem Identification

- Result not showing improvement as expected.
- Observation :

Training & Validation **Loss are high**

- Model might not fitting well and not able to learn pattern from training data.
- Therefore, **model architecture** might need to change.

#2 Feature Enhancement

Modification of Model Architecture

- Experiment with deeper architecture from original 64*64

Model Architect	Rationale	Total Profit	Sharpe Ratio	Max Drawdown
64*64*64	add additional layer while keeping the neuron count consistent, deepen the model	1.31	0.83	-0.31
256*256	Increases the neuron count in each layer, testing the impact of a wider network.	1.07	0.21	-0.39
128*128*128	balanced approach with more neurons per layer than the original and an additional layer, providing both depth and width	1.26	0.71	-0.33

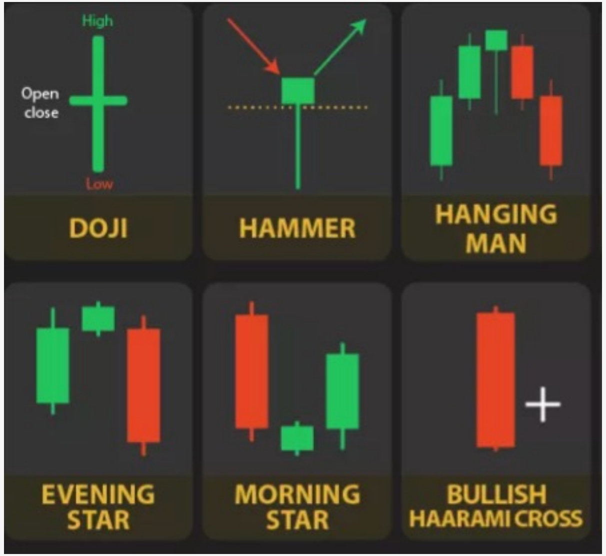
64*64*64 :

- Balanced Complexity
- Resource Efficiency
- Better Learning capability

#3 Feature Enhancement

Integrating Candlestick Pattern

- Candlestick pattern are pattern found by finance elites that used to predict next price movement.



Rationale

- Technical Indicators often lag the market, as it's calculated from past.
- Candlestick show future price movement

<u>Hypothesis :</u>

Candlestick = Better predicting.

#3 Feature Enhancement

Testing Result

Performance Metric	Results
Total Profit :	1.30
Sharpe Ratio :	1.11
Maximum Drawdown :	-0.15

Not improving much !

Discussion

- Reason why not performing well:

1. The prediction of the candlestick pattern is not necessarily true
2. Candlestick pattern data is usually **sparse**; not every time frame shows a clear pattern

- Still, remain potential as it has most numbers of features. At this stage, it might be more beneficial to shift focus to reward shaping (next phase of our study)

03
Methodology of Reward Shaping

Overview of Reward Shaping

Methodology

- Ori : Proportional Reward & Penalty
- #1 : Behavioral Reward
- #2 : Risk Handling via Thresholding
- #3 : Avoid Overconfidence

Original Reward Function

Proportional Reward & Penalty

- Direct reward and penalty based on the agent's action that leads to price difference of current trade and previous trade

```
price_diff = current_price - last_trade_price
step_reward += price_diff
```

Problem

- Too simple. Only cares about short-term gains from trading, no longer financial goals like long-term growth, risk management, or capital preservation.
- Doesn't have dynamic ways to change the rewards.

#1 Reward Shaping

Static Weight on losses

- Reward and penalty are treated differently
- Losses have a bigger effect on the person than gains (1.2 multiply to be exact)

Rationale

- Aligning with a principle of behavioural finance called **loss aversion**.
- RL agent will avoid losses rather than pursuing equivalent winnings

Hypothesis:

The agent can come up with a strategy that prioritise capital preservation first

#1 Reward Shaping

Testing Result

Performance Metric	Results
Total Profit :	1.47
Sharpe Ratio :	1.06
Maximum Drawdown :	-0.35

Improving

Discussion

- Result improving, but there’s one visible problem: static weights
- Therefore, doesn’t change as market condition change.
- Effort can be made on designing dynamic weights.

#1 Reward Shaping

Dynamic Weight on losses

- Penalty weight calculated by 20 days volatility.
- When market unstable (volatility high), penalty for losses goes up
- When market stable (volatility low), penalty is smaller.

Code Snippet

```
def calculate_normalized_volatility(prices, window=20):
    if len(prices) < window:
        return 1.0 # Default to no adjustment if not enough data
    recent_prices = prices[-window:]
    avg_price = np.mean(recent_prices)
    volatility = np.std(recent_prices) / avg_price # Normalized volatility as a
percentage of the average price
    return volatility
```

#1 Reward Shaping

Testing Result

Performance Metric	Results
Total Profit :	1.45
Sharpe Ratio :	1.35
Maximum Drawdown :	-0.21

Discussion

- Similar result as static weight, but much more reliable where:

1. Sharpe Ratio is higher
2. Max Drawdown is lower

We can continue develop reward function to multi objective like risk handling (#2 reward shaping)

#2 Reward Shaping

Risk Handling via thresholding

- Directly encourage our trading methods to keep risk under control
- Small reward for agent if stay above threshold, penalty if below.

```
step_reward = 0
if self._total_profit >= 1:
    step_reward += 1
else:
    step_reward -= 0.5
return step_reward
```

Testing Result

Performance Metric	Results
Total Profit :	0.91
Sharpe Ratio :	-0.36
Maximum Drawdown :	-0.53

#2 Reward Shaping

Problem

- Overall performance went down.
- Problem observe: Due to simple profit threshold and fix nature of weight.

Iterative Improvement

- Multiple efforts were made like trying different threshold, turning into dynamic weights.
- However, there is persistent of **bad performance** from reward shaping based on thresholding.
- Simplifying complex environment into **binary cases** might not be a good idea. We shall move to another angle

#3 Reward Shaping

Risk Handling via Avoid Overconfidence

- Look for new way to handle risk.
- Look into behavior of RL, problem of overtrading.
- Penalty agent if trading too much (3 trades in 10 window)

Rationale

- By controlling how often the trades happen, improve the model's general risk management.

Hypothesis :

Reduce possibilities of losses & Focus on high quality trades

#3 Reward Shaping

Testing Result

Performance Metric	Results
Total Profit :	1.69
Sharpe Ratio :	1.25
Maximum Drawdown :	-0.24

Discussion

- Very easy modification, but show significant result, which contrast with previous reward shaping (no need to have much complex reward function)
- The improved performance confirms our hypothesis.

04
Conclusion

4.1

Testing Result

Cross-Stock Testing

Rationale

To **validate** our methodology beyond the **original** Goldman Sachs (GS) stock dataset, we tested it on multiple stocks. This approach ensures that our strategies for feature enhancement and reward shaping are effective across **various market** conditions and stock types.

Stocks Selection

Diverse range of stocks to ensure the model handles various scenarios, including

APPL (Apple)	Technology
COST (Costco)	Retail
PFE (Pfizer)	Pharmaceuticals
KO (Cola)	Consumer Goods
XOM (ExxonMobil)	Energy

Same Training & Testing Process
Same Model configuration
Same Methodology
Same Evaluation

Stocks	Benchmark	Enhanced version
AAPL	Total profit : 0.97 Sharpe Ratio : -0.43 Max Drawdown : -0.12	Total profit : 1.45 Sharpe Ratio : 1.07 Max Drawdown : -0.17
PFE	Total profit : 0.74 Sharpe Ratio : -2.25 Max Drawdown : -0.26	Total profit : 0.83 Sharpe Ratio : -2.07 Max Drawdown : -0.25
COST	Total profit : 1.37 Sharpe Ratio : 3.66 Max Drawdown : -0.05	Total profit : 1.33 Sharpe Ratio : 3.55 Max Drawdown : -0.4
KO	Total profit : 0.98 Sharpe Ratio : -0.16 Max Drawdown : -0.12	Total profit : 1.12 Sharpe Ratio : -0.24 Max Drawdown : -0.18
XOM	Total profit : 0.94 Sharpe Ratio : -0.51 Max Drawdown : -0.19	Total profit : 1.41 Sharpe Ratio : 1.03 Max Drawdown : -0.28

Good

1. The improved models, integrating feature enhancement and reward shaping, **outperformed** the simple "buy and hold" benchmarks for most stocks
2. This success indicates the potential effectiveness of advanced RL methods in stock trading strategies.

Bad

1. Stocks like Costco showed **limited** improvement in overall profit and Sharpe ratio.
2. The methods were less effective in stable sectors with limited growth potential.

Possible Reason: Reward shaping focused more on **risk** management than on maximizing profit.

Solution : **Customizing** reward functions and hyperparameters for each stock based on its unique characteristics could further enhance results.

4.2

Summary of Work

Summary of Work

Feature Enhancement

Start with basic price data, to extended, to Technical Indicators, lastly candlestick pattern.

Reward Shaping

From Proportional reward and penalty, to static weight, to dynamic weight. Then, risk management with 2 approach, which is thresholding and avoid overconfidence.

Testing

Cross-stock validation to show the methodology is working.

What I've Learned

Application of
Reinforcement Learning

Basic knowledge of
Stock Trading

Importance of
Feature Enhancement

Reward Shaping as
double edge sword

Always learn from mistakes

Importance of
Iterative Improvement

Thanks!

Do you have any questions?

dc02615@umac.mo