Fin-search RoadMap 1

Investing in the stock market and maximizing the output is desire of every trader!

And to achieve these various types of techniques are used, like the technical indicators, automation by the machine learning. The optimising of the stocks portfolios and devising the effective investment strategies present the complex challenges that demand the innovative solutions.

In this roadmap we will explore the various characteristics of the traditional algorithms, including their manual requirements along with their limitations in the adapting to the ever-changing market conditions and the stocks prices. We will also dive in the reality of the RL-based algorithms, their adaptability, automation features and their potential for the scalable decision-making.

Via these two algorithms we can achieve some valuable insights of the RL that will help in the optimization and the investment strategies, leading the way for a self-learning which can navigate us through the various obstacles and the complexities of the financial markets journey.

Traditional algorithms:

In the context of the stocks market optimization and the investment strategies the various conventional methods that were applied to use the technical indicators in the finance before the era of the more sophisticated machine learning techniques. Before moving on the traditional algorithms let's just see the various technical indicators and their classification over which these algorithms are used.

Traditional Indicators are mainly categorized in the four categories viz;

Momentum-based

Trend-based

Volume-based

Volatility-based

Let's just take a brief about all of them along with the examples of the same and the strategies behind each of them.

Momentum-Based Indicators:

Momentum mainly refers to the capacity of any price trend to sustain itself ongoing be it the upward or the downward trend. Even more simply momentum is basically the inertia of the price trend to continue itself whether it is falling or rising trend for a particular amount of time and the indicators use the strength their momentum or the sustaining capability in determining the stocks movement direction.

Momentum indicators or MOM indicators are widely popular technical analysis tools used by traders to measure the rate at which the price of a stock fluctuates. They complement other indicators well as they only pinpoint a time in which the change in market price is taking place.

Let’s take the example of the momentum indicators:

The Relative Strength Index (RSI) and the Stochastic Oscillator are major example of the same and are most widely used.

Let’s not dive in the mathematical formulas and the calculation processes for now and focus on the technique behind it. As a momentum indicator RSI, it calculates the average gain and the average loss over a particular period and then calculates the RS ratio, which is further used to calculate RSI value, which ranges btw 0 to 100. More the RSI means the overbought situation and the vice-versa.

The stochastic oscillator compares the closing price to a range of prices over a certain period of the time. It relies on the stock price history and predicts the reversals if any. The sensitivity of the same depends on the period we have chosen

Trend-based:

As the name suggests these indicators see the trends in the stock prices and predicts if it’s going to repeat be it the uptrend, downtrend or the sideways.

MACD is the most famous example of the trend-based indicator which uses the exponential averages of various time periods. And the moving averages be it the simple one or the exponential averages all are based on the trend-based indicators.

Both uses the relative prices of the stocks over a particular period with giving more weightage to the recent ones in the exponential and determines the trend in them with the current prices.

Volume-based:

Volume, in short, we can say that it is the number of shares exchanged between buyers and sellers at any given interval. High volume provide liquidity and organized directional movement. Low volume generally results in unpredictable price movements with wide bid/ask spreads. This gives us the direction but not the magnitude of that direction that is why we look at the price movement and the volume to discern the difference.

This is important for us to have well known the volume as we the retailers are small fish in the market and generally, the institutions are the ones who make or break the market as per their will, so it is very important for us to have our trades aligned with the institutions so that we can make some profit as well.

Some examples of the volume-based indicators are the On-volume Based which add or subtract the closing prices as per the high or low. Also, the VWAP (volume weighted average) is also one of the examples of the same.

Volatility-based Indicators:

This is the range in which the prices can vary in any direction and hence they create a trading opportunity based on it. The faster prices change, the higher the volatility and the vice-versa.

The most common example of the volatility-based ones is the Bollinger bands. I am not going to explain the same as it is wider topic and it uses 3 bands, upper lower and the middle one.

Traditional Algorithms:

Now coming to the various traditional algorithms which are applied on these indicators to optimize the stocks when the automated RL strategies were not much dignified.

It won’t be possible to mention almost all the algorithms, but I will try to explain almost many of the famous ones!

1) Moving averages:
   This is the most common method where there are mainly two types of the averages are used viz the simple moving average (SMA) and the exponential moving average (EMA). As the name suggests the SMA is just the average of the stock prices over a particular period of the time. SMA formula is quite simple, so I am mentioning the EMA formula:-

EMA(t) = [Vt × m / (1+d)] + EMA(y) × [1 − (m/(1 + d)]

Where: EMA(t) = EMA today

Vt = Value today

EMA(y) = EMA yesterday

m = multiplier = 2 / (1 + d)

d = Number of days

The basic algorithm used by the people in the moving average is that short-term moving averages crossing above longer-term moving averages is generally seen as bullish and long-term moving averages crossing below short-term moving averages is generally seen as bearish.

2) Buy and Hold Strategy:

In this method it is just simple as a common man thinks what the trading is that is to buy when the prices fall and sell the stocks when the prices reach higher. This is generally a long-term investment.

3) Bollinger Bands:

In this algorithm the prices of the stock are quantified by the width of the upper and lower std deviation bands. Here too the basic layman method is applied traditionally that is it involves the use of three bands - one for the upper level, another for the lower level, and the third for the moving average. Prices moving toward the upper band are a sign that the market might be overbought. In contrast, if prices eventually drift towards the lower or bottom band, the market can be oversold.

Let's dive into the calculations, I am sticking here formulas from the internet sources for the band's calculation here

Bollinger bands are usually calculated based on a 20-period SMA of the candle close price :

$SMA_{20} = 1/20 \sum_{n = -19 \text{ to } 0} (close_n)$

The standard deviation is obtained according to the 20-period SMA on the same number of periods, so it is a moving standard deviation :

$$\sigma_{20} = \sqrt{\frac{\sum_{n=-19}^{0} (close_n - SMA_{20})^2}{20}}$$

The upper and lower bands are simply calculated by adding/subtracting twice the moving standard deviation to the 20-period SMA:

$Upper$ = $SMA_{20} + 2 \times \sigma_{20}$

$Lower$ = $SMA_{20} - 2 \times \sigma_{20}$

4) Markowitz's Modern Portfolio:
This algorithm mainly maximizes the returns for a given amount of the risk by maintaining a good relationship between the two. Here mainly the two lines are calculated Capital Market Line and the Capital Allocation Line based on the portfolio variance and std deviation. These lines provide a risk-free asset and reflect the risk tolerance along with providing the range of optimal portfolios.

I think in the technical indicators too I have discussed with some of the algorithms so let's move on the advantages and the disadvantages of these algorithms

Advantages:
Needless to mention there are lot of the advantages of these algorithms as they make us take fair and reasonable decisions not a game of the probability anymore.
Also, we can back test the previous data of the stock to see if our strategy is really giving us the maximized returns. And this also adds to the speed and efficiency which is not feasible for us to analyse such large amount of data of these stocks with their prices changing every time.

Apart from this we can combine two or more algorithms to make it accurate and precise in determination of the direction of the motion of the stock prices.

There are some laggings and the leading indicators which are useful to predetermine the stock movement and then the to confirm the trend respectively.

Disadvantages:
No matter these algorithms makes it simple for us to manage the prices but time to time changing these strategies according to the various stocks and for a various period of the time is quite hectic. And there is lot of the data dependency that is we need to collect and analyse the historic data of these stocks and calculate the indicators as per.

Also, the boundaries set by us on these indicators are not always accurate and are just the averages only. Risk management is also low in these algorithms and need to be modified a lot before applying to the real live market.

RL-Based Algorithms for STOCK trading.

Deep reinforcement learning algorithms can outperform human players in many challenging games. Return maximization as the trading goal: by defining the reward function as the change of the portfolio value, Deep Reinforcement Learning maximizes the portfolio value over time.

1) Deep Q Networks (DQN):

Deep Q-learning (DQN) and its enhancements are typically used to train a single agent on a specific stock or resource. The Critic alone approach focuses on learning the optimal action-selection strategy, which maximizes potential returns given the current situation, by utilizing Q-value functions. Instead of creating a state-action value table, DQN minimizes the error between estimated Q-values and target Q-values during training, employing a neural network for function approximation. This makes it suitable for continuous cost functions. However, the Critic alone approach has a significant drawback as it can only handle discrete and limited state and action spaces, making it impractical for scenarios involving a large number of stocks or resources. By addressing this limitation, the method becomes more efficient and applicable to broader domains.

$$Q(S_t, A_t) \leftarrow Q(S_t, A_t) + \alpha \left[\boxed{R_{t+1} + \gamma \max_a Q(S_{t+1}, a)} - Q(S_t, A_t) \right]$$

2) Deep Deterministic Policy (DDPG):

DDPG (Deep Deterministic Policy Gradient) has proven effective in maximizing investment returns . This algorithm combines elements of Q-learning and policy gradient methods, utilizing neural networks as function approximators. Unlike DQN, which suffers from the curse of dimensionality and relies on indirect adaptation through Q-value tables, DDPG directly benefits from feedback through policy gradients, making it more efficient.

The key idea in DDPG is to deterministically select actions based on states in an environment with a continuous action space. At each time step, the DDPG agent chooses an action □□ at state □□, receives a reward □□, and transitions to the next state

□□+1. These transitions (□□, □□,□□+1,□□) are stored in a replay buffer □. During training, a batch of □ samples is drawn from □, and the Q-value □□ for each sample is updated using the following formula:

$$y_i = r_i + \gamma Q'(s_{i+1}, \mu'(s_{i+1}|\theta^{\mu'}, \theta^{Q'})), i = 1, \cdots, N.$$

where □□ is the reward at time step □, □ is the discount factor, and θ_μ' and θ_θ' are the target network parameters. The actor network's output, denoted by □(□□|θ_μ), represents the action chosen by the actor, and the critic network's output □□(□□, □□|θ_θ) estimates the Q-value for the state-action pair.

By updating the critic network using the temporal difference error and optimizing the actor network's parameters to maximize the expected return, DDPG learns a deterministic policy that maps states to actions, leading to improved investment returns. The combination of Q-learning and policy gradients in DDPG makes it a powerful algorithm for continuous action spaces and a suitable choice for investment optimization tasks.

In DDPG, the critic network is updated by minimizing the loss function, which corresponds to the predicted difference between the outputs of the target critic network Q' and the current critic network Q.

$$L(\theta^Q) = \mathbb{E}_{s_t, a_t, r_t, s_{t+1} \sim \text{buffer}}[(y_i - Q(s_t, a_t|\theta^Q))^2].$$

3) Advantage Actor Critic (A2C):

A2C is a normal actor-critic calculation & we use it part in ensemble technique. It is a type of the reinforcement learning where the agent mainly focuses on the actor and the critic. A2C uses benefit capacity through diminish difference of strategy gradient. Rather than just estimates value capacity, critic network assesses benefit work. A2C is an incredible model for stock exchanging on account of its stability. Target work for A2C is :

$$\nabla J_\theta(\theta) = \mathbb{E}[\sum_{t=1}^{T} \nabla_\theta \log \pi_\theta(a_t|s_t) A(s_t, a_t)],$$

where □□ (□□ |□□) is policy network, □(□□, □□) is Advantage function can be written as:

$$A(s_t, a_t) = Q(s_t, a_t) - V(s_t),$$

or

$$A(s_t, a_t) = r(s_t, a_t, s_{t+1}) + \gamma V(s_{t+1}) - V(s_t).$$

It maintains a good balance between the simplicity of the policy gradient methods and the stability of the value-based methods.

4) Proximal Policy Optimization (PPO):

PPO is that art of the reinforcement learning algorithms that mainly focuses on the limitation of the traditional gradient system. Policy gradient methods use stochastic policies and rely on sampling trajectories to estimate the gradient of the expected reward with respect to the policy parameters. This method controls the arrangement gradient update it & ensures that new approach isn't too far from previous one.

Let us assume that probability proportion between old & new approaches is given as:

$$r_t(\theta) = \frac{\pi_\theta(a_t|s_t)}{\pi_{\theta_{old}}(a_t|s_t)}$$

The clipped & conventional goals are used through calculate PPO [19]'s target capacity. Outside of clipped span, PPO discourages major strategic shifts. And, hence as a result, by limiting arrangement update at each preparation step, PPO improves strength of strategy network preparation. Here we choose PPO for stock trading since it is more stable, rapid, & simple through use.