A Portfolio Management Strategy Using MDP Model

Group 39 ,Xie Long

SIGS, Tsinghua University, China

Abstract

• **Develop a data mining pipeline to partition the states for the changing process of stock price in A-share**

• **Design a transformed markov decision model to represent the environment and manange the portfolio**

Motivation

The interpretability of most portfolio management strategies is poor

The application of Markov decision model based on discrete state in portfolio management is rarely attempted

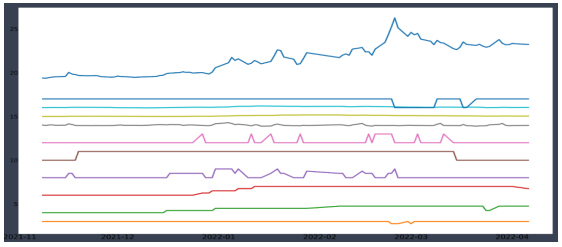
Most of the factors used in traditional portfolio management strategies are ineffective and cannot provide special heterogeneous returns

Traditional strategies rarely take into account the positive feedback of the self strengthening phenomenon of the stock market

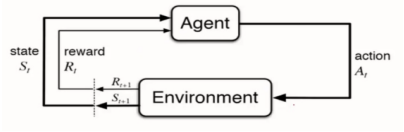
In this project, we extracted some key features, constructed some key factors, and designated the states for each trading day , and developed a MDM+Transformer based portfolio management strategy

• Preliminary preparation

We extract about 20 synthesis factors, as shown in the figure(part of them), which will be used in the process of states partioning.



Method



• We initially define the Dynamic Observable Markov Decision Model(DOMOM) by the tuple:

$$(S,\ P,\ A,\ R,\ \pi(s,\ r),V,\Omega,\beta)$$

1).To better model the strategy, our model partitions the changing process of price each day into 7 different discret state.

2).The state transition probability P is dynamicly calculated.As time goes by, P keeps being updated.

$$P_{ij} = 1/4P_{ijself} + 3/4P_{ijmarket}$$

$$P_{ijX} = \frac{P(S_{t+1} = S_j, S_t = S_i)}{P(S_t = S_i)}, (X = self, market)$$

3). A is an action sets, which describe buy and sell operations.

4).The reward function R calculates the expected the Sharp Ratio for each state, which depends on the current state, the next state and the probability of state transition.

5).π(s,r) is a policy function whose input is (s, r) and output is A(the action sets), where (s, r) represents (state, reward) and A represents action for state S. Policy function determine action a that should be selected under state S.

6).Vπ(s) is the overall evaluation for a specific state, which reflects the expected value of the expected cumulative return under state s.

$$V_\pi(s) = E[G_t|S_t = s]$$

$$Q_\pi(s,a) = E[G_t|S_t = S, A_t = A]$$

$$Q_\pi(s,a) = \sum_{s' \in S} P^a_{s \to s'}(R^a_{s \to s'} + \gamma V(s'))$$

$$\forall s \in S, V^*(s) = V_{\pi^*}(s)$$

$$V^*(s) = \max_{a \in A} Q^*_\pi(s,a)$$

$$\pi^*(s) = arg\max_{a \in A} Q^*_\pi(s,a)$$

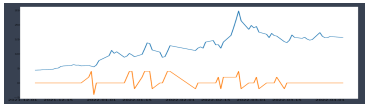
7).Omega is a set of observation rules,which is determined by us.

8) . Beta is a value to evaluate the market systematic risk.

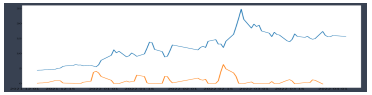
Results

(1).State partitioning

The above curve is the daily price, and the following is the corresponding states.

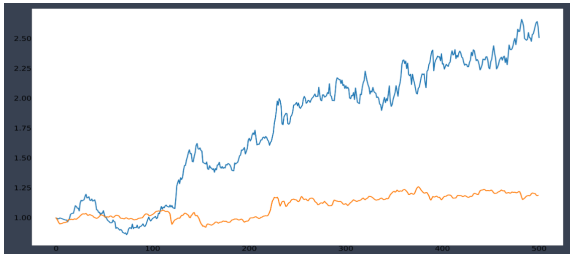


(2).Overall value of state



(3). Performance of model and strategy

The above curve(blue) is the Cumulative Net Return Curve of our model in 500 tradingdays, and the following (yellow) is the cumulative net return curve of CSI 1000 in the same period.



method	Annualized Return	Sharpe Ratio	Maximum rollback
ours	**75%**	**3.01**	**23%**
CSI 1000	**6%**	**1.08**	**28%**

Conclusion

1).The influence of impact cost on results is not considered , so the capacity of portfolio management strategy is relatively small.

2).Because too much personal experience is used, the results are actually over fitting.

3).There is little interaction with the environment and the State transition is not a markov process, so the model is not a MDM literally .

References

[1] A. Gouberman and M. Siegle, “Markov reward models and markov decision processes in discrete and continuous time: Performance evaluation and optimization,” in International Autumn School on Rigorous Dependability Analysis Using Model Checking Techniques for Stochastic Systems. Springer, 2012, pp. 156 – 241.