Improving Reinforcement Learning Trading Model with CDC ATR Trailing Stop Strategy

Thiti Leelasomphop[1], Ekarat Rattagan[2]

[1]Graduate Students at Business Analytics and Data Science, Graduate School of Applied Statistics National Institute of Development Administration

[2]Assistant Professor Dr. at Business Analytics and Data Science, Graduate School of Applied Statistics National Institute of Development Administration

Corresponding author: Thiti Leelasomphop (e-mail: thiti.lee@stu.nida.ac.th).

ABSTRACT For newbie human traders, one of the easiest ways to profit from the market is by following trade signals from expert-created strategies that have already been proven with historical data. The CDC ATR Trailing Stop strategy, invented by Mr. Chaloke Sambandaraksa, a well-known technical trader in Thailand, suggests when traders should buy or sell, and can also be used as a trend indicator. We consider this information as a suggestion from the expert and it may be able to improve reinforcement learning models. Therefore, in this paper, we propose a method for incorporating the information from the CDC ATR Trailing Stop strategy into a reinforcement learning trading model to make it profitable. Specifically, we add the suggestion from the strategy to the reward function as a penalty component. If the model opens a position in the opposite direction from the suggestion from the strategy, it will be punished. With this penalty component, the agent will try to make a profit while avoiding trading in the opposite direction of the suggestion. We use this method to train and test our model on historical gold price data and found that the added penalty component has a significant impact on the model's profitability. Our proposed model outperforms both the original CDC ATR Trailing Stop strategy and the reinforcement-only model that does not adopt this penalty component. The model not only generates higher profit but also provides a better profit to maximum drawdown ratio when compared to the mentioned baseline models in every testing period from 2018 to 2021.

INDEX TERMS Reinforcement Learning, CDC ATR Trailing Stop, Financial Trading, Machine Learning

I. INTRODUCTION

People have been interested in trading for a very long time since it is one of the ways to financial freedom. People can make profit from the market by buying and selling the securities at the proper time. However, with high volatility, deciding when to buy or sell is always difficult, especially for the newbie trader. The technical analysis seems to be the technique which require a personal experience. One of the methods which famous among newbie trader is following the signal from some trading strategy which some experts created and validated with historical data to make profit. The strategy guides the trader when they should and should not buy or sell. This kind of trading strategy was not mentioned so much in academic terms but well known among technical traders. CDC ATR Trailing Stop Strategy was founded by Mr.Chaloke Sambandaraksa, famous technical trader of Thailand. The strategy tells the trader whether price of securities is currently in buy zone or sell zone. With this strategy, newbie traders can follow the signal to open and close their positions without analyzing the technical indicators themselves. Anyway, even if it was validated with historical data, there are no guarantee that this system can generate profit in every period of time. Sometimes, we can find the false signal which lead to loss or the profit which lower than our expectation due to its too fast or too slow response. The information given from the strategy is very useful but there still have a room for improvement to get better trading strategy.

Recently, the machine learning techniques has been used in many businesses, including the investment. One of the algorithms which being very popular is the reinforcement learning technique which best fit with timeseries data like the securities 's price. There are many studies that try using reinforcement learning technique to find the profitable strategy. In this paper our aim is to find the profitable trading strategy using reinforcement learning technique. But instead of letting the agent find out the strategy itself, we guide it to trade following CDC ATR Trailing Stop strategy.

II. BACKGROUND

A. Technical Indicator & Trading Strategy

A technical indicator is a mathematical calculation based on historic price and volume, that aims to forecast financial market direction.

A trading strategy or trading system is basically a trading plan with a set of rules that define the entry and exit conditions to go in and out of the market.

B. CDC ATR Trailing Stop Strategy

CDC ATR Trailing Stop is an indicator created by "Chaloke-Dot-Com" (www.chaloke.com) which founded by Mr.Chaloke Sambandaraksa, one of Thai famous technical trader. This indicator is the combination of the technical indicator named "Average True Range" and the "Trailing Stop" technique.

Average True Range is an indicator which calculate the average of a "True Range" defined in the formula below.

$$TR = Max[(H\text{-}L), Abs(H\text{-}C_p), Abs(L\text{-}C_p)]$$

$$ATR = \left(\frac{1}{n}\right)\sum_{(i=1)}^{(n)} TR_i$$

Where :

H = *High price of the day*

L = *Low price of the day*

C_p = *Close price of previous day*

TR_i = *A particular true range*

n = *The time period employed*

The trailing stop technique is a modification of typical stop order that can be set to close the opened positions at some amount away from current market price. If the security price rises or falls in your favor, the stop price moves with it. If the price rises or falls against you, the stop stays in place.

When the price is uptrend, this indicator will be plotted under the price with green color. Whenever the price is drop down under this green line, the color will be change to red and the indicator will move to be plotted above the price. This indicator alone can be set as a trading strategy itself. As mentioned earlier, the buy signal is given when the indicator turns from red to green and the sell signal is given when it turns from green to red.

C. Reinforcement Learning

Reinforcement Learning (RL) is a type of machine learning technique which has an agent inside the environment. Agent will choose an action based on the observation that he sees. Then the environment will change and feedback a reward to an agent. Agent will learn from the reward that he received to improve his action selecting in the future. This learning concept is really proper for training a trading model. We can let the securities market be an environment and the agent be an investor. The agent will use securities price and some technical indicators to decide the action to buy, sell or hold the securities. Then he will receive the profit or loss as a reward.

III. RELATED WORKS

Lure V. Brandao et al.[4] had adopted some famous trading strategy, i.e., RSI overbought/oversold, EMA cross over, MACD cross over, etc., to train reinforcement learning model. The trading signal from these strategies were used to assign the reward for trading agent. If the agent giving a trade signal at the same time and same direction as the signal from these strategies, it will receive the positive reward. If the agent gives a trade signal in the direction opposite to signal from the strategies it will be punished with negative reward value.

Joao Carapuco et. al.[2] had made a reinforcement learning trading model for a forex market. Some interesting techniques were used to avoid overfitting and improving model performance. Instead of training the model with whole data each time, the researcher made an environment which will randomly pick a period of data to train an agent on each iteration. With this method the model will see different data on each training iteration. This method made the model be able to maintain its profitability even if on unseen data. In addition, the researcher split some data just before each test data to be used as validation data. While training, the model was evaluated on the validation data set regularly. The reward for each evaluation will be observed. After training ended, the model which provide best result in validation data set will be continually used for testing on test data set. This model be able to generate profit stably during test data from 2010 to 2016.

Lin Li[3] had proposed to do some feature preprocessing on the data before training. The Principal Component Analysis (PCA) and Discreate Wavelet Transform (DWT) were studied. The researcher chose 11 famous technical indicators. Then used the mentioned preprocessing technique to these indicators before using them to train the reinforcement learning agent. The result showing that both techniques be able to improve the trading performance of an agent.

Xing Wua et al.[1] compared 2 reinforcement algorithms for trading, Q-learning and Policy gradient. The result showing that both algorithms have good performance. However, the policy gradient algorithm, which is Actor-Critic type, be able to generate more stable return than the Q-learning, which is Critic only type.

IV. EXPERIMENTAL SETUP

A. Data & Features

The historical of daily gold price from 2008-2021 including Open, High, Low, Close and Volume was used.

From original price data, we created 11 financial indicators. The indicators data will be normalized by Z-Score method. Then using "Principal Component Analysis" (PCA) technique to reduce the dimension while the output components still contain 95% of explained variance. Finally, the 11 features were reduced to 8 components.

B. Reinforcement Learning

Environment

Our trading environment is created by the library called "Open AI Gym". In this environment, we restrict the agent to be able to hold only one position at a time. If he wants to open new position, the previous one need to be closed before that. The position size is set to be always 0.01. Change of gold price for 1 USD will result in our profit or loss for 1 USD. The price which agent use to open position is always the close price of that trading day. To make the environment simple, we set the trading fee to always be 0.3 USD for both opening and closing the position instead of using actual spread.

Agents

In this study, we try using 3 algorithms from library called 'Stablebaseline3'. One is a Q-learning algorithm, Deep Q-Learning (DQN). The other 2 algorithm are Policy Gradient, Proximal Policy Optimization(PPO) and Asynchronous Advantage Actor Critic (A2C).

Actions

Agent has 3 possible actions to choose which are "Buy", "Hold" and "Sell". Each action will give different result depend on the environment at that time. Here is where we adopt the "CDC ATR Trailing Stop" into our work.

FIGURE 1. Sample data for CDC ATR Trailing Stop Strategy and our proposed method of defining buy zone and sell zone

From original CDC ATR Trailing Stop Strategy, the investor should "Buy" the security when the indicator turns from red to green. Then "Hold" it whenever the indicator still being green and "Sell" the bought security when indicator turn into red. The short-sell position will do this process in opposite. We got some idea from this strategy. It suggests to keep "Buy" position as long as the indicator being green and keep "Sell" position as long as it being red. This may be able to conclude that when the indicator is green, the possibility that the price will continue rising up is higher and it better to keep buying. In other word, selling when indicator is green may has higher risk. We think that we can guide our RL agent using this idea. Therefore, we set up one parameter called "Penalty" and add it to reward function. We define the area which the indicator is being green as a "Buy Zone" or green zone. And the area which the indicator is being red as a "Sell zone" or red zone. Opening sell position in green zone or opening buy position in red zone are considered as a risky trade. If our agent doing this, we will punish it with some amount of penalty. The penalty value is set to be constant and considered as one of hyperparameter which need to be tuned.

TABLE I

Action	Open position	Zone	Result	Penalty
0	-	BUY	Do nothing	None
		SELL		
	BUY	BUY		
		SELL		
	SELL	BUY		
		SELL		
1	-	BUY	Open BUY position	None
		SELL		Penalty add
	BUY	BUY	Do nothing	None
		SELL		
	SELL	BUY	Close SELL position	
		SELL		
2	-	BUY	Open SELL position	Penalty add
		SELL		None
	BUY	BUY	Close BUY position	
		SELL		
	SELL	BUY	Do nothing	
		SELL		

Reward

We separate our reward function into two parts. The first part is the difference between current net worth and net worth of last timestep. This part will give a positive reward to agent when it be able to increase our total asset value and punish it with negative value when our asset value is decrease. The other part is the penalty part which already mentioned before. The reward function can be written as shown below.

$$Reward = Net\,Worth_i - Net\,Worth_{i-1} - Penalty$$

Where :

Net Worth $_i$ = Total asset value at time step *i*

Penalty = Penalty value depended on agent's action

With this reward function, we expect to see our agent trying to grow our net worth value while avoiding to open the position in the opposite way from CDC ATR Trailing Stop Strategy. Anyway, we guess that the agent will try to trade against the strategy sometimes if it thinks that the total accumulate return in the future is worth to get this amount of penalty.

Observation

The observation data that we use to train the model is composed of the financial indicators that already reduced via PCA technique, the difference of close price of current day and previous, number of position that agent currently hold, mean and standard deviation of close price difference, color of zone (green or red) and the distance of current price to stop loss line from CDC ATR Trailing Stop Strategy.

TABLE II
OBSERVATION

Features	Number of lookback days	Data size
8 Components of reduced financial indicators	5 days	40
Difference of Close price between current day and the day before	5 days	5
Number of position that agent currently hold	1 day	1
Mean of Close price different between current day and the day before	1 day	1
Standard deviation of Close price different between current day and the day before	1 day	1
Zone indicated by CDC ATR Trailing Stop Strategy (Green/Red)	1 day	1
Distance between price of current day to the stop loss line of CDC ATR Trailing Stop Trailing Strategy	1 day	1

Training & Testing

We used the time-based cross validation data technique. The data since 2008 are used as train data set and data of 2018, 2019, 2020 and 2021 are used as test data set. In addition, we used 2 techniques to avoid overfitting.

First, instead of using whole training set of data and let our agent see same observations on every iteration of training, we randomly pick some periods of data to train on each iteration. Therefore, our agent sees different data at each time. This will make our model more generalize and having better trading performance on unseen data.

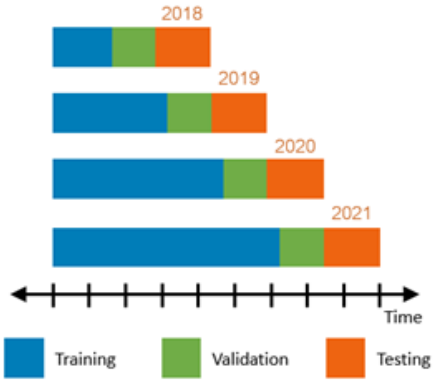

FIGURE 2. Schema of data sets divisions for validating and testing

Another technique is separating some data at just before the test data as validation data. Normally, if we keep continue training the model, it will continue getting better rewards. Anyway, at one point, this model will start to be overfit and giving bad performance on unseen data. We use the validation data set to solve this problem. At first, we set our target number of training timesteps to be a large number(N). Then we set another small number for evaluation interval(n). While training, at every n timesteps, we evaluated our model at that time on validation data set and recorded it. After finished training N timesteps, we picked the model that provide best performance on validation data to continue testing on test data set.

Hyperparameters

Since there are too many parameters that need to be tuned, optimizing is very time consumed. We try tunning them with only some attractive numbers. Finally, the parameters which doing best are listed in table III. Parameters which not specified in this table are all set to be default values from Stablebaseline3 library.

TABLE III
HYPERPARAMETERS

Symbol	Description	Value
WS	Training window size	750
VS	Validation data size	250
N	Target training timesteps	2,000,000
n	Training evaluation interval	2,500
P	Penalty	20
LR	Learning rate	0.001

Metrics

The metrics which we use in this study are : Net profit which directly show the profitability of the model on each of testing period in USD; Maximum Drawdown which measures the net worth declination from peak value before largest drop and lowest value before new high established during each test period; The ratio between Net profit and Maximum Drawdown, the risk-adjusted return metric. We expect to see high Net profit with low Maximum Drawdown which indicate that our model has high profitability while be able to manage risk well.

In addition, there are another 2 metrics which not be able to represent the model performance directly but we want to measure to know the model behaviors. These 2 are “Number of trades” and “Average gain per trade”. These two metrics represent how often the model doing trading activities and also its strategy, whether get small gain but getting frequently or getting large gain but only few times of trading.

Baselines

In this paper, we compare the proposed trading model with 2 baselines: the CDC ATR Trailing Stop Strategy, which giving signal for buying and selling every time the color of indicator changes; the Reinforcement Learning only model, which use all same data and method as our proposed model except the penalty component (penalty always be zero). For the Reinforcement Learning only model, we use the algorithm which doing best in overall trading period.

V. EXPERIMENTAL RESULT

The numerical metrics show in table IV. Among 3 agents we used in this study, we found that PPO can generate highest profit. Therefore, we create the “Reinforcement Learning Only model” using PPO algorithm, to be used as a baseline.

The result shows that our proposed model can beat both original CDC ATR Trailing Stop Strategy and the RL-only model in all testing period from 2018-2021. Comparing profit, our proposed model is the best, follow by CDC ATR Trailing Stop Strategy and RL-only model consequently. The Penalty component has really high impact to our model. The RL-only model which doesn’t has penalty component can only make profit in 2020 while our proposed model not only beat the baselines but also be able to generate profit stably in every testing year. Even if in 2021 which gold price swing in small frame. Original CDC ATR Trailing Stop strategy faces many false trading signals and lose the money but our model still fine and be able to keep making profit. For the Maximum drawdown, our model also doing well. It has lower value than other 2 baselines. These lead to better result in profit to maximum drawdown ratio which we can say that our model has higher profit ability than the baselines, comparing at same risk level.

Checked at trading activities of our model on test data, we found some interesting behaviors. The model always tries to hold a position, either buy or sell. Because it wants to maximize its accumulative reward, so it needs to invest. We also found the behavior that we expected. The model sometimes picks a risky trade, open buy position on red zone and open sell position on green zone. This let our model be able to get higher profit since it can open the position earlier when price trend is change. However, we found many times that our model picks risky trades and failed. But the model manages these trades quiet well. After open the risky position, if the price is not going as expected, model usually close these positions shortly, before it’s going

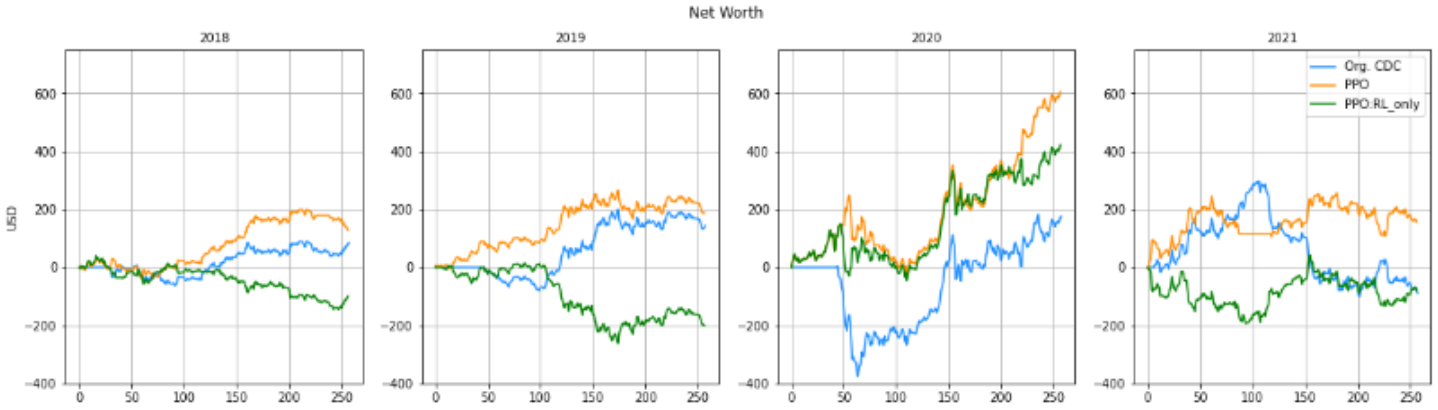

FIGURE 3. **Net worth value comparing our proposed model to original CDC ATR Trailing Stop Strategy and Reinforcement Learning only model**

worse. This behavior well known among experience investors as "cutting lost" which we think this is the main reason why our model can get lower maximum drawdown value.

TABLE IV
NUMERICAL RESULTS

		2018	2019	2020	2021	Total
Profit (USD)	PPO	128.0	187.3	604.9	156.5	1076.7
	Org. CDC	82.8	143.8	174.6	-89.2	311.9
	RL only	-99.5	-201.3	422.0	-78.8	42.6
	A2C	-93.7	-234.8	142.5	-391.4	-577.4
	DQN	39.0	-162.2	820.4	317.2	1014.4
Max DD (USD)	PPO	-73.6	-98.9	-273.6	-148.6	-
	Org. CDC	-70.4	-87.4	-379.9	-396.7	-
	RL only	-185.5	-277.1	-195.3	-193.4	-
	A2C	-145.1	-327.9	-337.3	-434.8	-
	DQN	-106.9	-180.8	-162.8	-224.1	-
Profit : Max DD	PPO	1.7	1.9	2.2	1.1	-
	Org. CDC	1.2	1.6	0.5	-0.2	-
	RL only	-0.5	-0.7	2.2	-0.4	-
	A2C	-0.6	-0.7	0.4	-0.9	-
	DQN	0.4	-0.9	5.0	1.4	-
Number of trades (times)	PPO	4.5	9.5	13.5	19.5	49
	Org. CDC	5.5	5.5	7.5	8.5	29
	RL only	27.5	10.5	13.5	29.5	83
	A2C	2.5	3.5	3.5	5.5	17
	DQN	0.5	2.5	18.5	19.5	43
Average gain (USD/time)	PPO	25.6	18.7	43.2	7.8	22.0
	Org. CDC	13.8	24.0	21.8	-9.9	10.8
	RL only	-3.6	-18.3	30.1	-2.6	0.5
	A2C	-31.2	-58.7	35.6	-65.2	-34.0
	DQN	39.0	-54.1	43.2	15.9	23.6

VI. CONCLUSION

In conclusion, our proposed model using CDC ATR Trailing Stop Strategy can significantly improve the profitability of the reinforcement learning. Our proposed model not only provide higher profit comparing to CDC ATR Trailing Stop strategy and RL-only model but also be able to generate stable profit for all 4 testing periods, from 2018-2021. Furthermore, the proposed model also better in risk management. It can generate higher profit comparing at same risk level.

VII. FUTURE WORK

As previously mentioned, the reinforcement learning model usually sensitive to hyperparameter tuning. Different combinations of parameters may have high impact on model performance. Unfortunately, since there are too much parameters, we had tried only few sets of them in this study. Therefore, optimizing the hyperparameters for this model is one of the challenge tasks which we plan to do in the future. In addition, we know that we can guide the reinforcement learning agent via the reward function. In this study, we used only CDC ATR Trailing Stop strategy to punish the agent when it selects a risky action. In the future, it is interesting to punish the agent with other consideration. It is possible to add other trading strategy or use technical analyst's opinion to consider when should we punish the model.

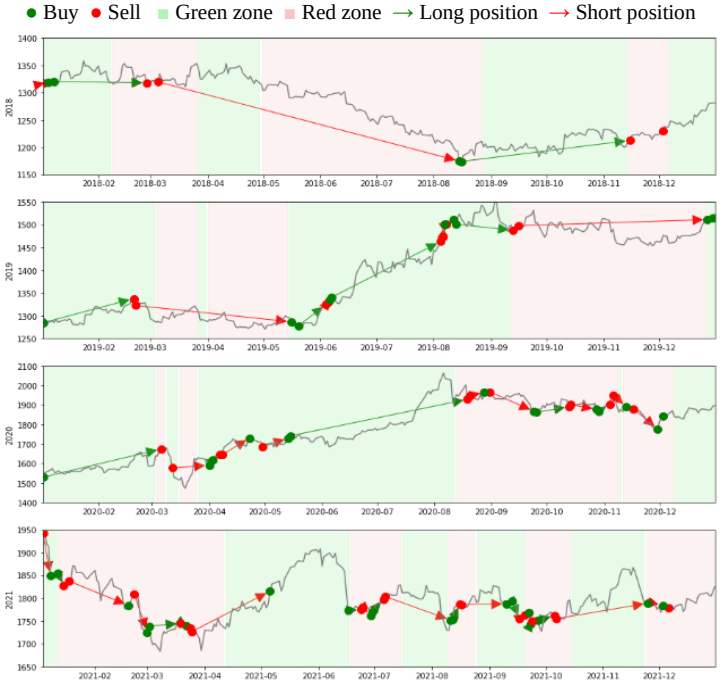

FIGURE 4. **Trading behaviors of our proposed model on testing period, from 2018 to 2021.**

VIII. REFERENCE

[1] "Adaptive stock trading strategies with deep reinforcement learning methods", Xing Wua, Haolei Chen, Jianjia Wang, Luigi Troiano, Vincenzo Loia, Hamido Fujita

[2] "Reinforcement Learning Applied to Forex Trading", João Carapuço, Rui Neves, Nuno Horta

[3] "Financial Trading with Feature Preprocessing and Recurrent Reinforcement Learning", Lin Li

[4] “Decision support framework for the stock market”, Iure V. Brandao, Joao Paulo C. L. da Costa, Bruno J. G. Praciano, Rafael T. de Sousa Jr., Fabio L. L. de Mendonca
[5] “Technical analysis strategy optimization using a machine learning approach in stock market indices”, Jordan Ayala, Miguel García-Torres, José Luis Vázquez Noguera, Francisco Gómez-Vela, Federico Divina
[6] “A method for automatic stock trading combining technical analysis and nearest neighbor classification”, Lamartine Almeida Teixeira, Adriano Lorena Inácio de Oliveira
[7] https://stable-baselines3.readthedocs.io/en/master/#
[8] https://www.chaloke.com/forums/topic/cdc-atr-trailing-stop/
[9] https://www.gymlibrary.ml/
[10] https://araffin.github.io/post/sb3/