

Improving Reinforcement Learning Trading Model with CDC Trailing Stop Strategy

Thiti Leelasomphop

6310422006

BA & I

Introduction .

Overview : Trading Strategy



_____ ●

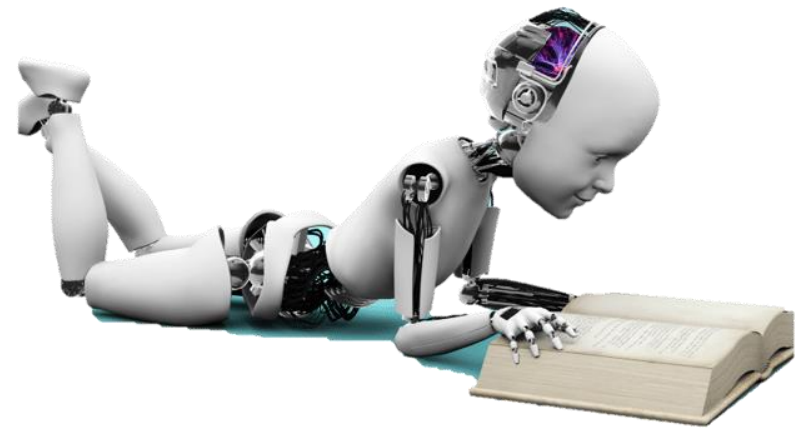


Motivation

—.



Technical Trading Strategies



Machine Learning

Background .

Technical indicator & Trading strategy.

Technical indicator

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

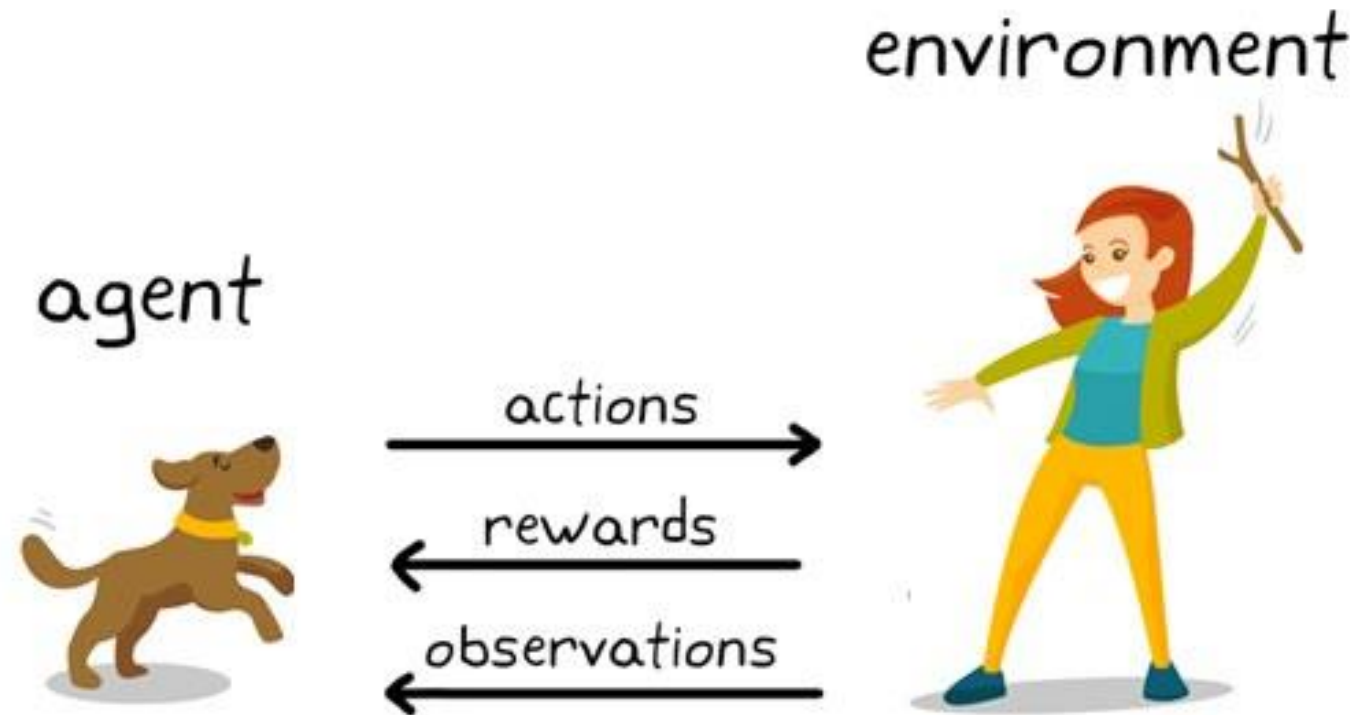
Trading Strategy (Trading system)

- A 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

Trading Strategy : CDC ATR Trailing stop



Reinforcement Learning .



Reinforcement Learning .



Literature review

No.	Title	Authors	Securities	Comments
1	Decision support framework for the stock market using deep reinforcement learning	Iure V. Brandao, Joao Paulo C. L. da Costa, Bruno J. G. Praciano, Rafael T. de Sousa Jr., Fabio L. L. de Mendonca	USA & Thailand stock market	<ul style="list-style-type: none"> - นำสัญญาณซื้อขายที่ได้จาก technical strategies มาสร้างเป็น feature - Train reinforcement learning model โดยนำโมเดลให้สัญญาณซื้อขายตรงกับ technical strategies ก็ให้รางวัล ถ้าให้สัญญาณซื้อขายที่ตรงกันข้ามก็จะลงโทษ - โมเดลที่ได้ให้ผลตอบแทนที่ดีกว่างานวิจัยก่อนหน้านี้ที่ใช้ NN
2	Reinforcement Learning Applied to Forex Trading	João Carapuço, Rui Neves, Nuno Horta	EUR/USD	<ul style="list-style-type: none"> - นำ reinforcement learning model มาใช้ในการเทรดระยะสั้น (tf : tick data) - ในการ train model จะทำการสุ่มจุดเริ่มต้นของข้อมูลที่ใช้เป็น observation เพื่อให้โมเดลมีความ generalize มากขึ้น - มีการแบ่งข้อมูลส่วนที่ใกล้กับ test data ออกมาส่วนหนึ่งเพื่อใช้สำหรับการ validation โดยจะนำโมเดลที่ให้ผลลัพธ์ดีที่สุดในช่วง validation ไปใช้ในการทดสอบกับ test data - จากการแบ่ง test data ตั้งแต่ปี 2010-2016 โมเดลสามารถทำกำไรได้ 18 จาก 21 ช่วงที่ทำการทดสอบ
3	Financial Trading with Feature Preprocessing and Recurrent Reinforcement Learning	Lin Li	<ul style="list-style-type: none"> - NYSE idnex - Exxon mobile coporation stock - Corn future contract 	<ul style="list-style-type: none"> - นำ technical indicator 11 ตัวมาทำ feature processing ด้วย PCA และ Discrete wavelet transform ก่อนที่จะนำข้อมูลนี้ไปใช้ train ด้วย reinforcement learning model - จากผลการศึกษาพบว่า PCA และ DWT ทำให้โมเดลมีประสิทธิภาพที่ดีขึ้น โมเดลให้ผลตอบแทนที่ดีกว่าการ trade โดยใช้กลยุทธ์ Buy&Hold, ARIMA และ โมเดล RL ที่ไม่ได้ทำ feature processing
4	Adaptive stock trading strategies with deep reinforcement learning methods	Xing Wua, Haolei Chen, Jianjia Wang, Luigi Troiano, Vincenzo Loia, Hamido Fujita	Stocks from <ul style="list-style-type: none"> - US - UK - China 	<ul style="list-style-type: none"> - ใช้เทคนิค Gated recurrent unit มาทำการ process ข้อมูลก่อนส่งเข้าไป Train ใน RL model - ใช้ Model RL 2 แบบ คือ Deep Q-learning (Critic only) และ Policy gradient (Actor-Critic) - โมเดลทั้ง 2 แบบให้ผลตอบแทนที่ดีกว่า Technical strategy (Advance turtle trading system) แต่โมเดลที่เป็น Actor-Critic ให้อัตราผลตอบแทนที่เสถียรกว่าโมเดลที่เป็น Critic only

Problem Statement .

- How to train a reinforcement learning agent to be able to make profit from the market ?

Solution .

- Improve the reinforcement learning model using technical trading strategy
 - Add 'Penalty' component to reward function when RL agent do some action that technical trading strategy suggest not to do

Experimental Setup .

Data Source .

- Data : Historical gold price since 2008 (tf :daily)
- Source : Meta Trader5

	Date	Open	High	Low	Close	Volume
0	2008-01-02	834.98	861.25	834.25	857.35	17840
1	2008-01-03	857.60	869.38	853.23	864.72	22597
2	2008-01-04	864.80	868.85	854.30	859.25	19707
3	2008-01-07	859.75	864.88	855.13	857.80	19863
4	2008-01-08	857.85	881.10	856.75	877.74	21310

Reinforcement Learning - Agent .

- Library : Stable-Baselines3
 - Proximal Policy Optimization (PPO)
 - Asynchronous Advantage Actor Critic (A2C)
 - Deep Q Network (DQN)

Reinforcement Learning - Environment .

- Library : OpenAi Gym
- Max number of hold position : 1
- Lot size : 0.01
- Trading fee : 0.3\$
- Open position price : 'Close' price of that trading day

Features Engineering .

- Features Processing
 - Technical indicators
 1. MACD : Moving Average Convergence/Divergence
 2. MOM : Momentum indicator
 3. MFI : Money Flow Index
 4. RSI : Relative Strength Index
 5. ATR : Average True Range
 6. NATR : Normalized Average True Range
 7. HTDCP : Hilbert Transform Dominant Cycle Phase
 8. SINE : Hilbert Transform Sinewave
 9. HTTM : Hilbert Transform Trend vs Cycle Mode
 10. CO : Chaikin Oscillator
 11. OBV : On Balance Volume
 - Dimension reduction
 - Principle Component Analysis → ~95% explained variance
 - Technical indicator : 11 → 8 components

Trading Strategy : CDC ATR Trailing stop



Reinforcement Learning - Action

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		

Will be added
to Reward



Reinforcement Learning - Reward

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

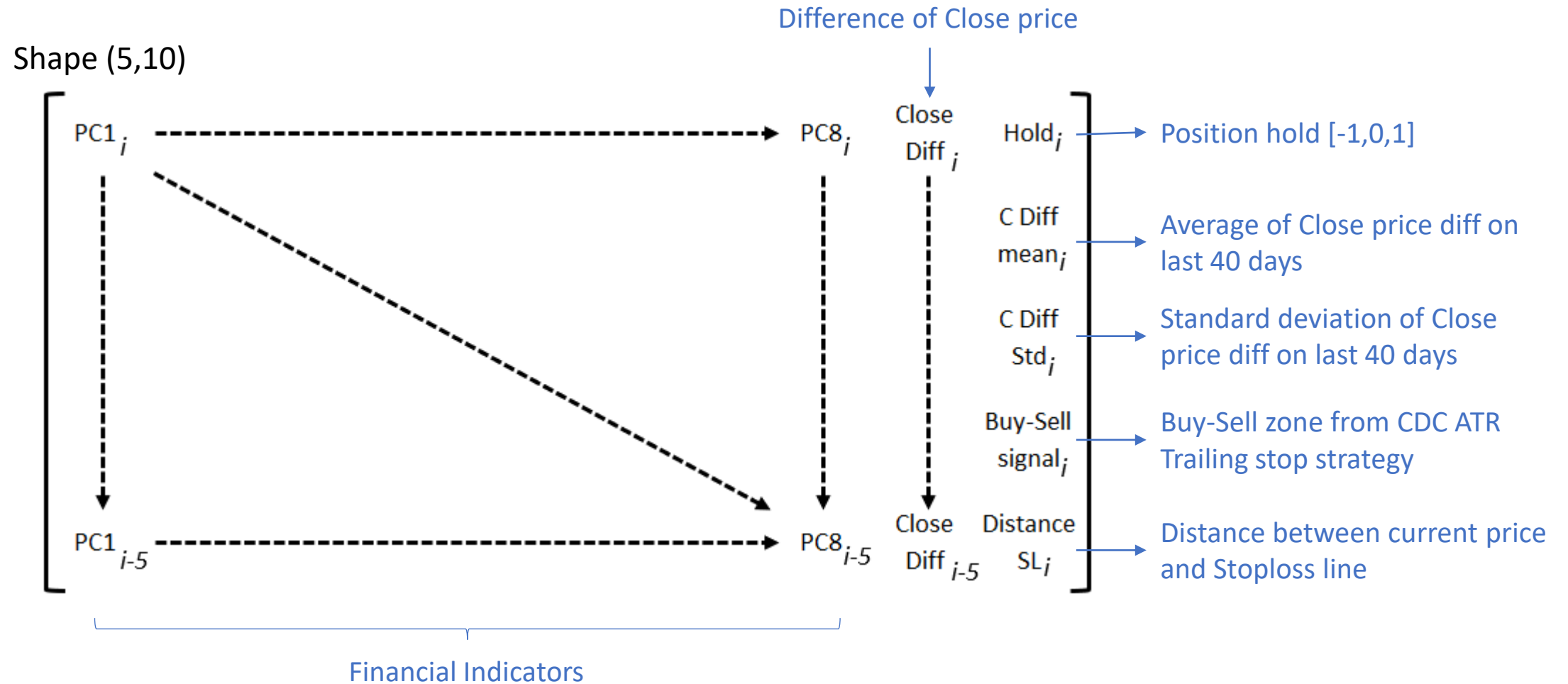
Where :

Net Worth_i = Cash + Current gain of open position at time step i

Net Worth_{i-1} = Cash + Current gain of open position at time step i - 1

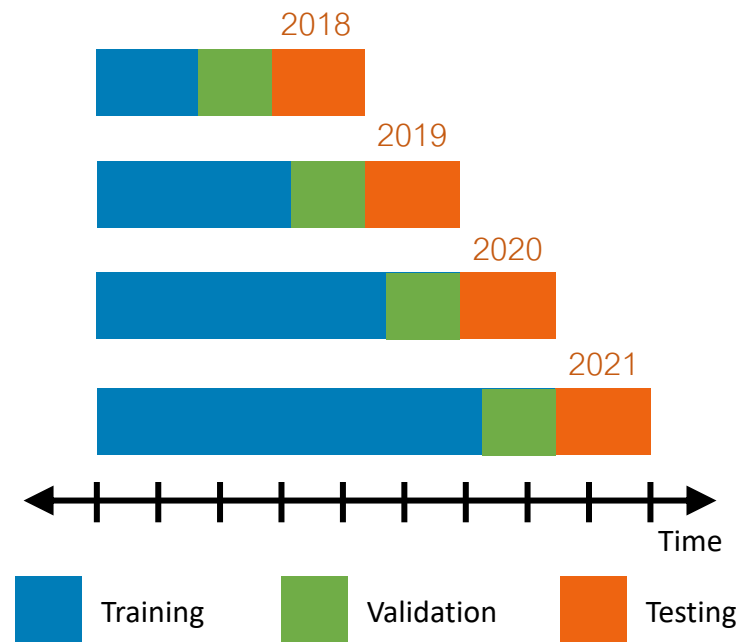
Penalty = Penalty value, depend on Action and Market condition

Reinforcement Learning - Observation



Train-Test Data .

- Time based cross validation



Train-Test Data



Hyperparameter.

- Training window size : 750 trading days
- Validation size : last 250 trading days
- Training timesteps : 2,000,000 timesteps
- Model Evaluation Interval : every 2,500 timesteps
- Penalty : 20 USD

Metrics.

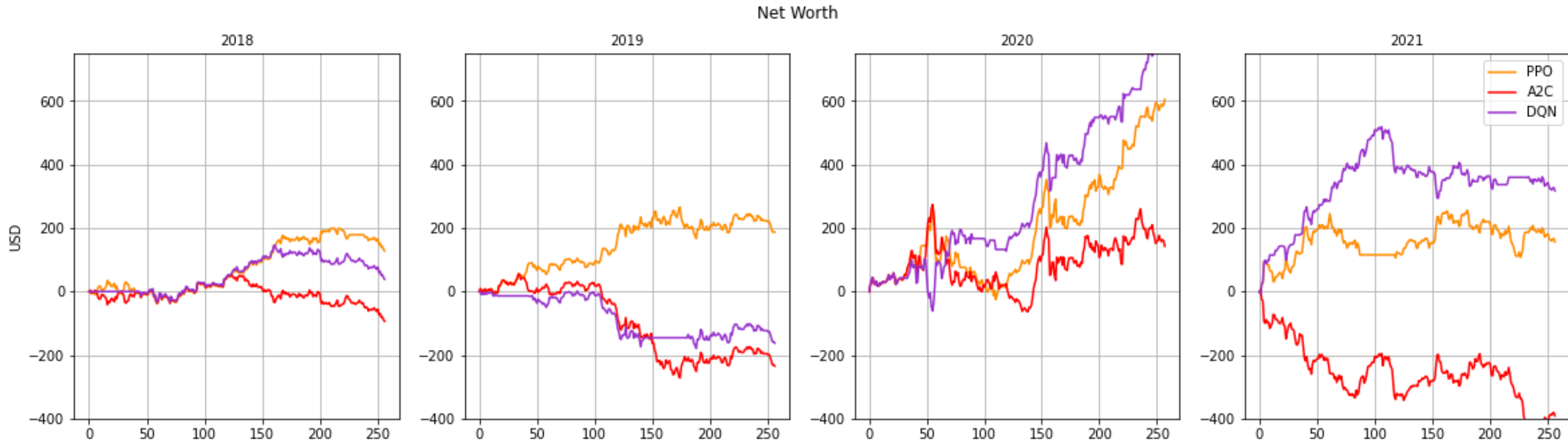
- Profit
- Maximum drawdown (Net worth maximum decrease)
- Profit to Maximum drawdown ratio
- Number of trades
- Average gain per trade

Benchmark.

- Original trading strategy : CDC ATR Trailing Stop
- 'RL only' model : Model without Penalty component

Experimental Result .

Result : Net worth



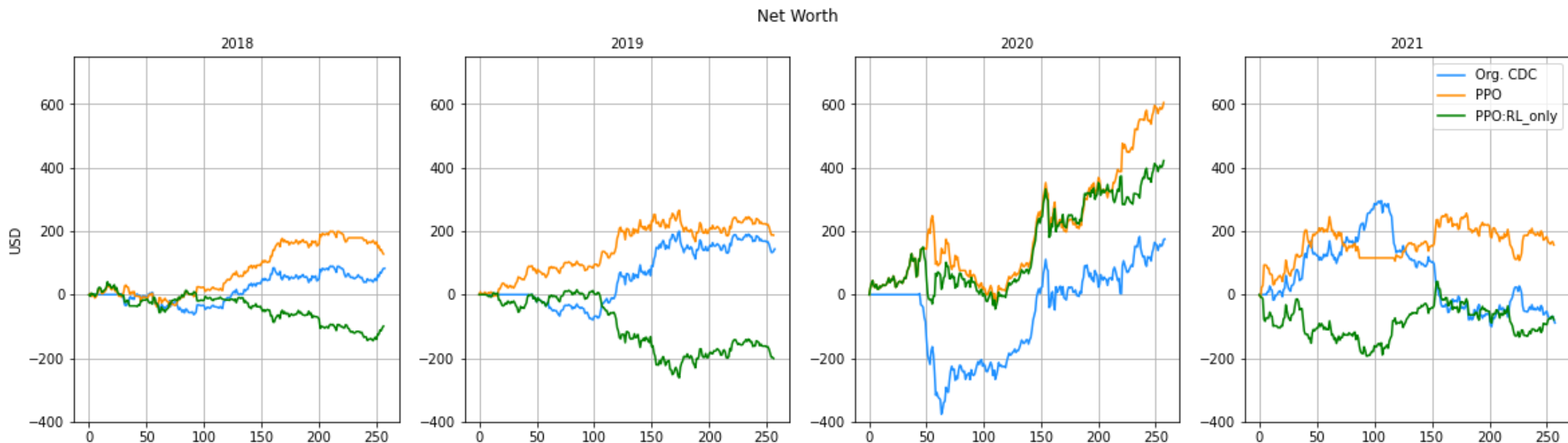
Total Profit :

PPO = 1,076.7 USD

A2C = -577.4 USD

DQN = 1,014.4 USD

Result : Net worth

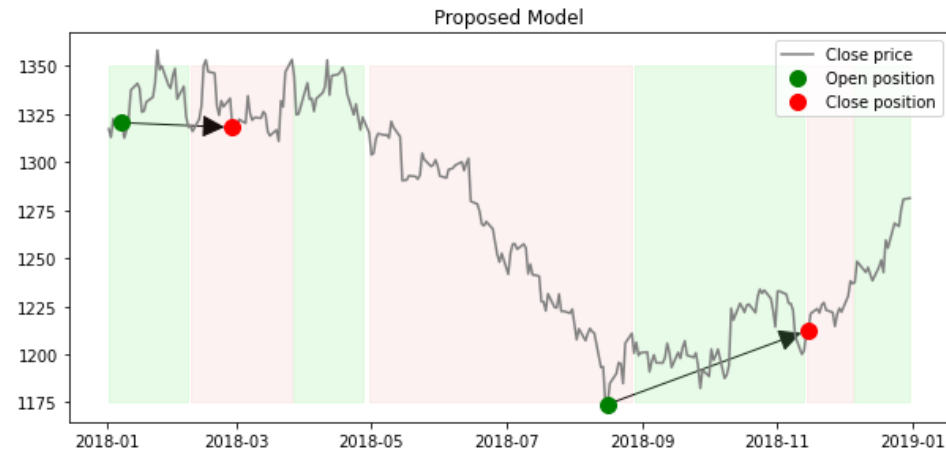


Result : Summary

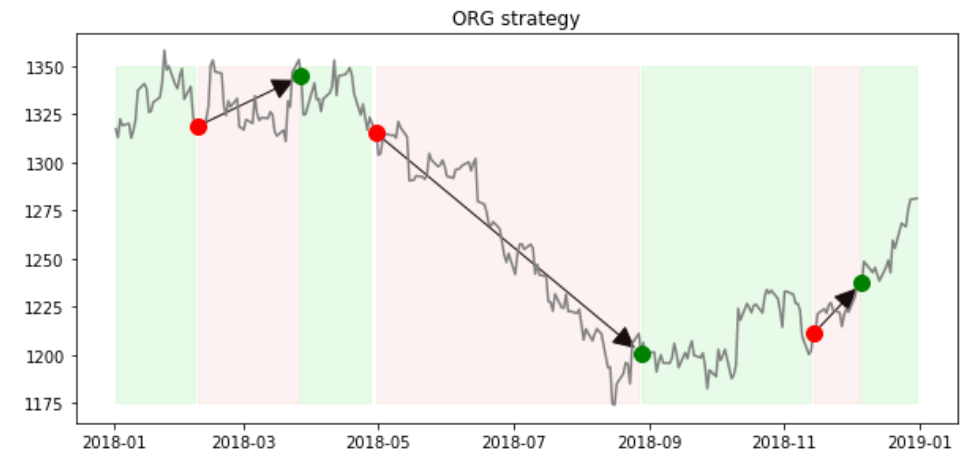
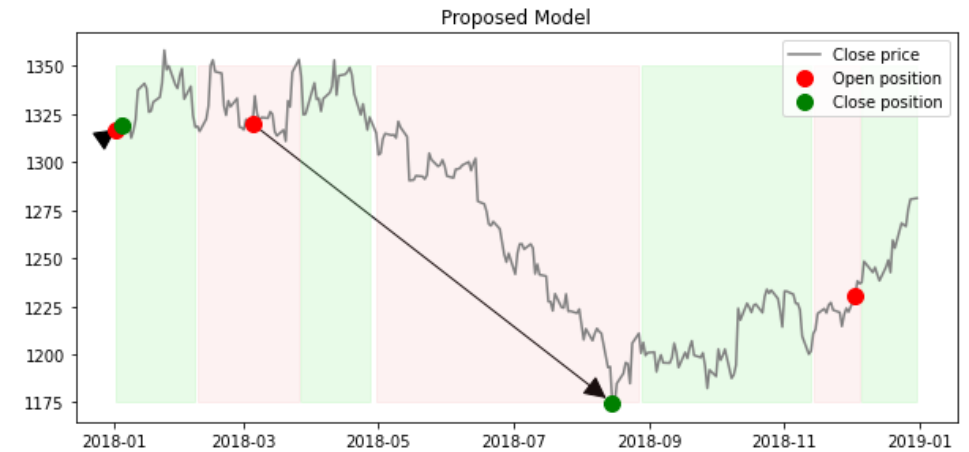
		2018	2019	2020	2021	Total
Profit (USD)	PPO	128.0	187.3	604.9	156.5	1076.7
	Org. CDC	82.8	143.8	174.6	-89.2	311.9
	RL only	-99.5	-201.3	422.0	-78.8	42.6
	A2C	-93.7	-234.8	142.5	-391.4	-577.4
	DQN	39.0	-162.2	820.4	317.2	1014.4
Max DD (USD)	PPO	-73.6	-98.9	-273.6	-148.6	-
	Org. CDC	-70.4	-87.4	-379.9	-396.7	-
	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

Result : Trading point (PPO model:2018)

Long trade

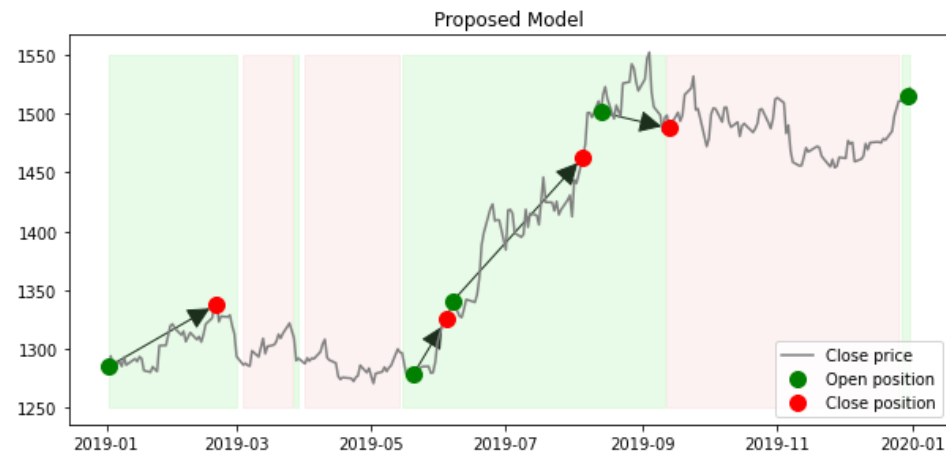


Short trade

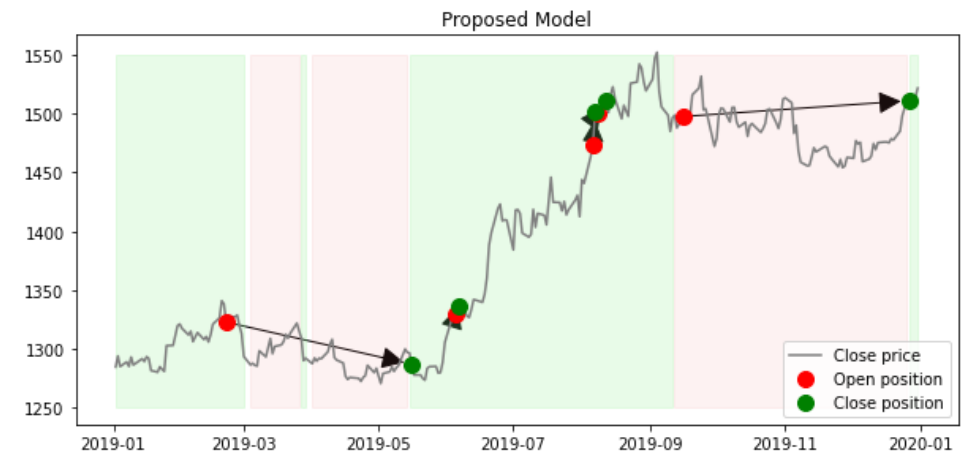


Result : Trading point (PPO model:2019)

Long trade



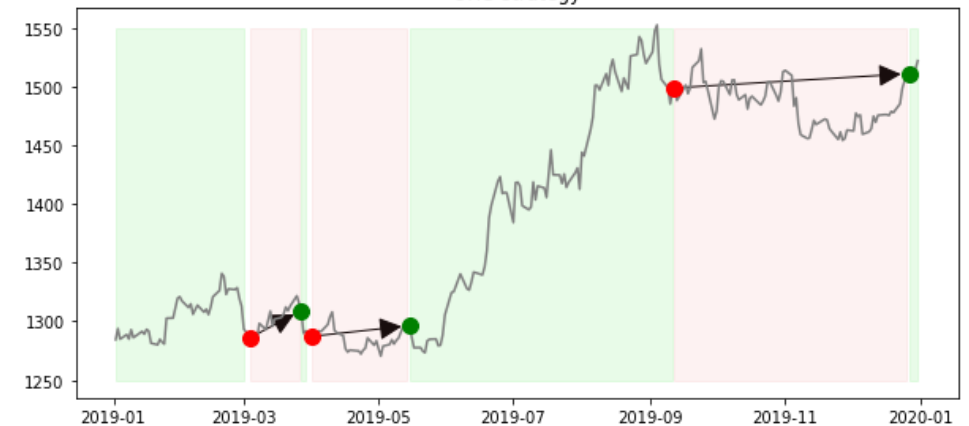
Short trade



ORG strategy

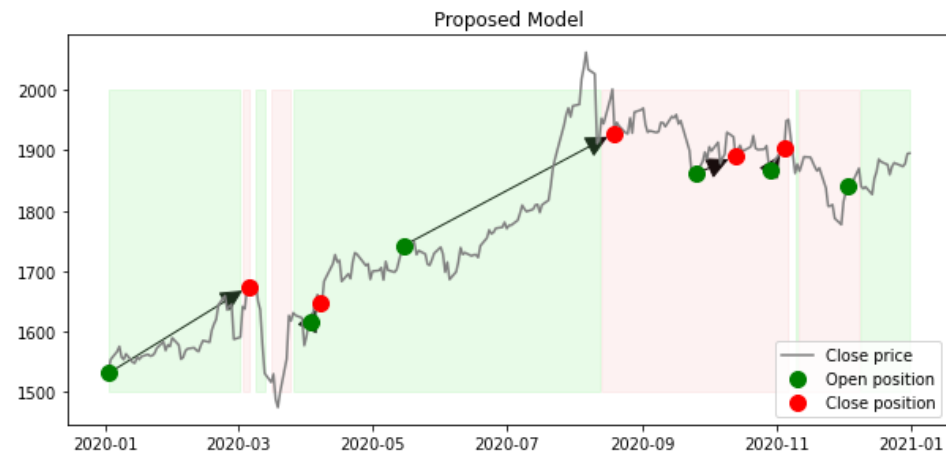


ORG strategy

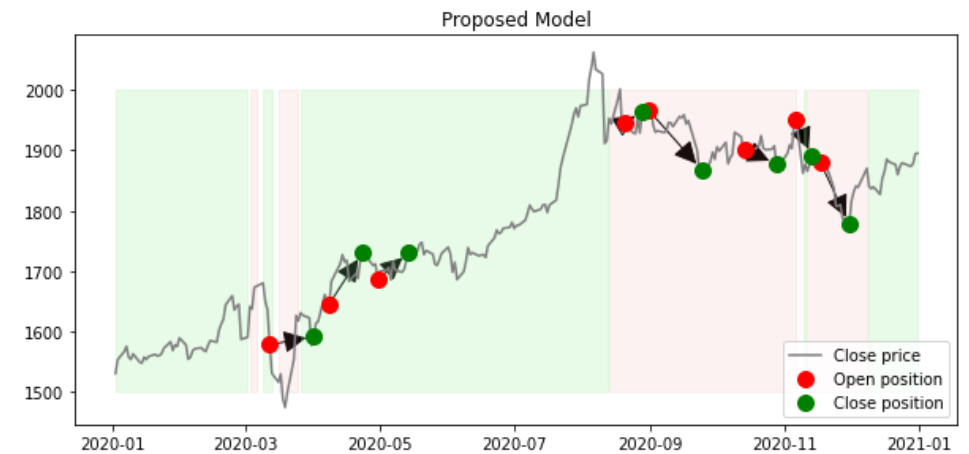


Result : Trading point (PPO model:2020)

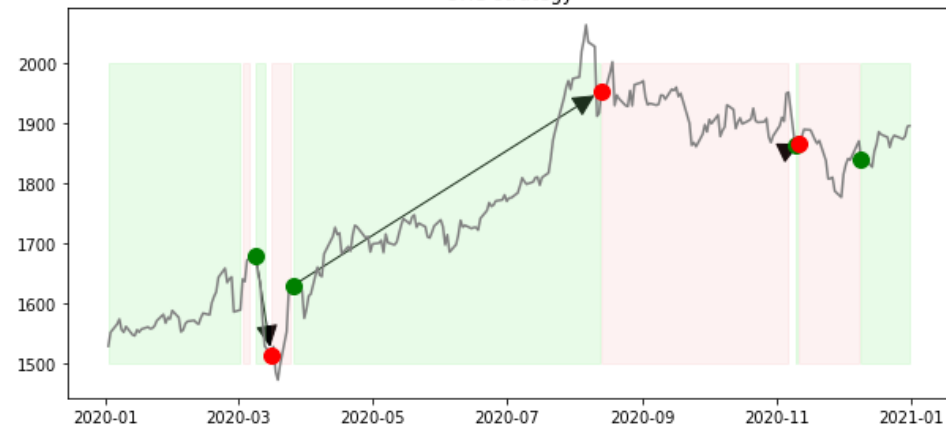
Long trade



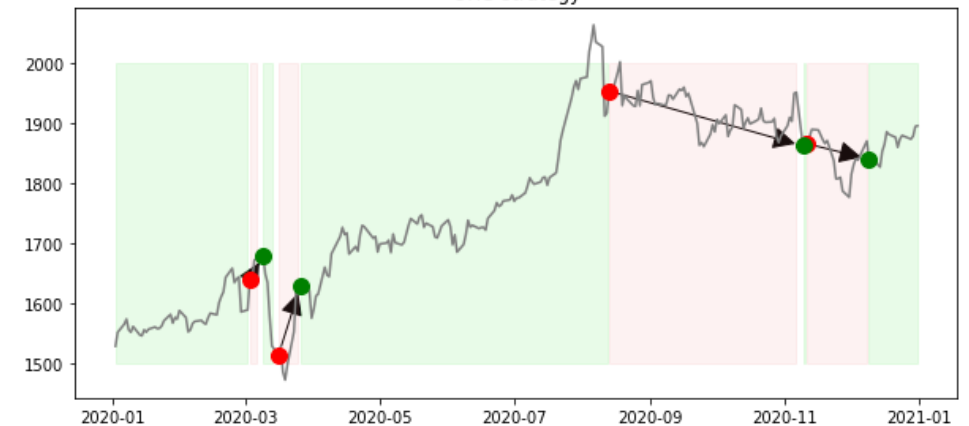
Short trade



ORG strategy

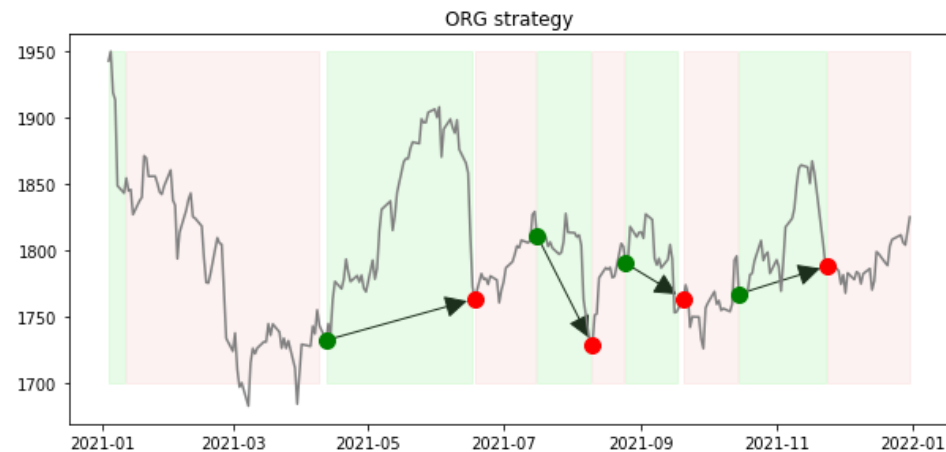
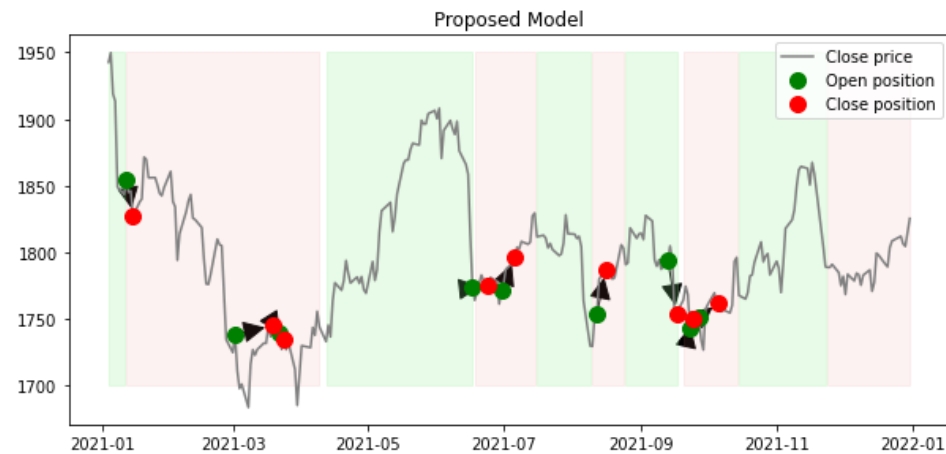


ORG strategy

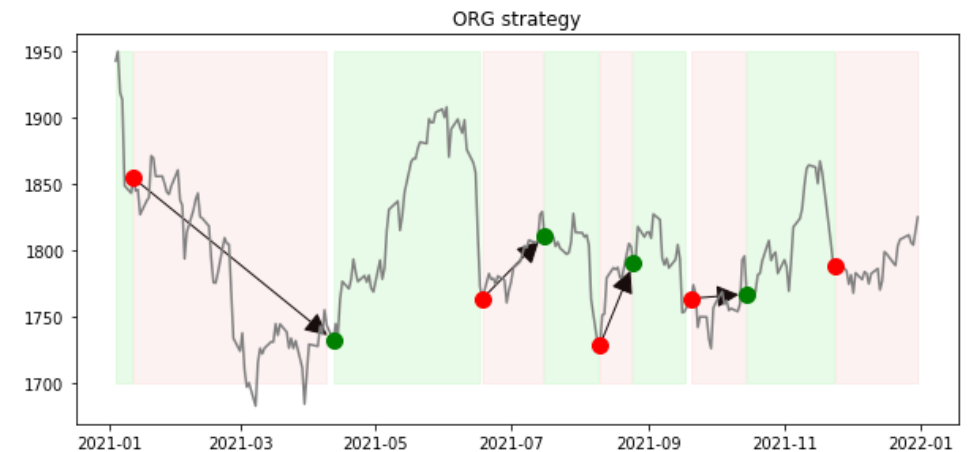
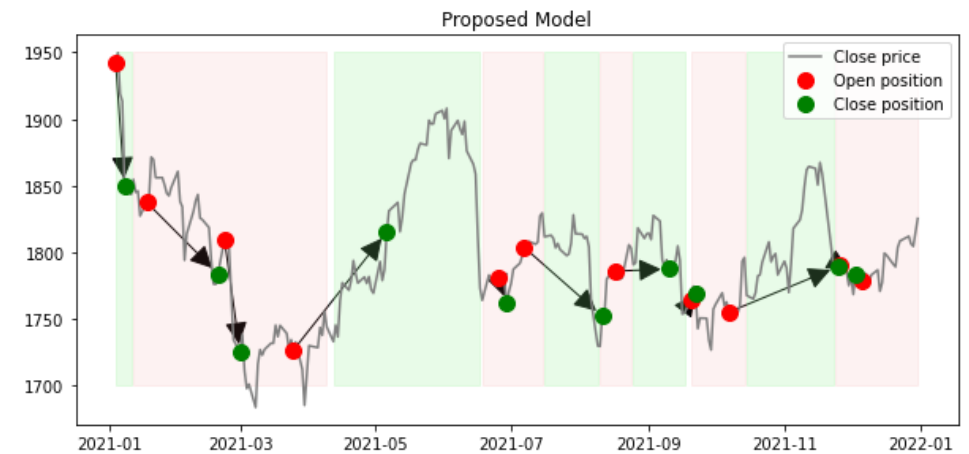


Result : Trading point (PPO model:2021)

Long trade



Short trade



Conclusion.

- Adopting CDC ATR Trailing Stop Strategy into the reward function as a penalty component can significantly improve the profit ability of the reinforcement learning model.
- The proposed model can generate higher profit comparing to both CDC ATR Trailing Stop original strategy and the reinforcement model without penalty adoption.
- The model with PPO agent can generate highest profit and be able to make profit in every single year of testing data, from 2018 to 2021.
- Comparing to CDC ATR Trailing Stop system, our proposed model has higher potential to generate profit, comparing at equal risk level.

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/>

Thank you.