Deep Reinforcement Learning for Trading

21. 11. 15

Hyogeun Park

Outline

- Introduction
- Methodology
- Experiments
- Conclusion

Introduction

- **Financial trading has been a widely researched topic and a variety of methods have been proposed to trade markets over the last few decades.**
  - **Fundamental analysis**
  - **Technical analysis**
  - **Algorithmic trading**
  - **Hybrid of these techniques**

- **Algorithmic trading has arguably gained most recent interest and accounts for about 75% of trading widespread, ranging from strong computing foundations, faster execution and risk diversification.**

Introduction

- **Limits of Algorithmic trading**
  - Low signal-to-noise ratio of financial data
  - Low signal-to-noise ratio of dynamic nature of markets
  - The design of these methods is non-trivial and the effectiveness of commonly derived signals varies through time

- In recent years, machine learning algorithms have gained much popularity in many areas, with notable successes in diverse application domains
  - Most research focuses on regression and classification pipelines in which excess returns, or market movements, are predictive signals into actual trade positions

Introduction

- In this paper…
  - Reporting on RL algorithms to tackle some problems
    I. Transforming these predictive signals into actual trade positions
    II. Need a signal with good predictive power but also a signal that can consistently produce good directional calls
  - Adopting state-of-art RL algorithms
    I. Deep Q-learning (DQN)
    II. Policy Gradients (PG)
    III. Advanced Actor-Critic(A2C)

Methodology

- State Space
  - Price
  - Returns
  - MACD(Moving Average convergence Divergence)
  - RSI(Relative Strength Index)

- Price
  - Normalize close price series

- Returns
  - Returns over past month, 2-month, 3-month and 1-years periods are used

Methodology

- **Returns**
  - **Returns over past month, 2-month, 3-month and 1-years periods are used**
  - **Normalize annual returns as**

$$\text{return} = \frac{r_{t-\ period,t}}{s_t\sqrt{period}}$$

where

$r_{t-\ period,t}$: $period$-day return

s_t : exponetially weighted moving standard deviation of r_t with 60day span

$$s_t = \left((1-\ l\ )r_{t-1}^{\ 2} + l\ s_{t-1}^{\ 2}\right)^{\frac{1}{2}}$$

Methodology

- **MACD(Moving Average Convergence Divergence)**

$$\mathrm{MACD}_t = \frac{q_t}{\mathrm{stdev}\left(q_{t-\ 252:t}\right)}$$

$$q_t = \frac{EWMA(S) - EWMA(L)}{\mathrm{stdev}\left(p_{t-\ 63:t}\right)}$$

where

p_t : price at time t

stdev : standard deviation

S : short time scale(in here =(8, 16, 32))

L : Llong time scale(in here =(24, 48, 96))

$$EWMA(x_t) = l\ EWMA(x_{t-\ 1}) + (1 - l\ )x_t$$

where

$$l = \frac{n - 1}{n},\ n : \text{time scale length}$$

Appendix A

- **EWMA(Exponentially Weighted Moving Average)**

$$V_t = l\,V_{t-1} + (1-l)Q_t$$

Q_t : New input data or present sample

$$V_t = l\,V_{t-1} + (1-l)Q_t$$

$$V_{t-1} = l\,V_{t-2} + (1-l)Q_{t-1}$$

$$V_t = l\,(l\,V_{t-2} + (1-l)Q_{t-1}) + (1-l)Q_t$$

$$= l^2 V_{t-2} + l\,(1-l)Q_{t-1} + (1-l)Q_t$$

$$= l^3 V_{t-3} + l^2(1-l)Q_{t-2} + l\,(1-l)Q_{t-1} + (1-l)Q_t$$

Methodology

- MACD(Moving Average Convergence Divergence)

Methodology

- **RSI(Relative Strength Index)**
  - **Oscillating indicator moving between 0 and 100**
  - **Indicating the oversold(<20) or overbought(>80) conditions of and asset by measuring the magnitude of recent price changes**

$$\mathrm{RSI} = \frac{\mathrm{RS}}{(1+\mathrm{RS})}$$

$$\mathrm{RS} = \frac{\mathrm{ave}(U)}{\mathrm{ave}(D)}$$

$$U: p_t - p_{t-1} \ (\text{if } p_t - p_{t-1} > 0)$$

$$D: p_t - p_{t-1} \ (\text{if } p_t - p_{t-1} < 0)$$

Methodology

- **Action Space**
  - **For discrete action spaces**

$$A_t = \{-1, 0, 1\}$$

I. **Each values represents the position directly**
II. **-1 mean maximally short position**
III. **0 mean no holdings**
IV. **1 mean maximally long position**

  - **For continuous action spaces**

$$A_t = [-1, 1]$$

Note.

1. If the current action and next action are the same, **no transaction cost** will occur and we just maintain previous positions
2. If we move from a fully long position to a short position, **transaction cost will be doubled**

Methodology

- **Reward Function**

$$R_t = m_{t-1} \frac{s_{tgt}}{s_{t-1}} r_t - \text{bp} ? p_{t-1} \left| \frac{s_{tgt}}{s_{t-1}} A_{t-1} \quad \frac{s_{tgt}}{s_{t-2}} A_{t-2} \right|$$

where

s_{tgt} : volatility target(constant)

s_t : exponentially weighted moving standard deviation with 60-day window on r_t

r_t : return at the time $t (= p_t - p_{t-1})$

bp : basis point(constant, for transaction cost)

$$s_t = \left((1 - l) r_{t-1}^2 + l s_{t-1}^2 \right)^{\frac{1}{2}}$$

Experiments

- Use data on 50 ratio-adjusted continuous futures contracts from the Pinnacle Data Corp CLC Database(2005-2019)
- Variety of asset classes : commodity, equity index, fixed income and FX
- Retrain the model at every 5 years, using all data available up to that point to optimize the parameters
- Model parameters are then fixed for the next 5 years to produce out-of-sample results
- Testing period : 2011 ~ 2019

Experiments

- **Baseline Algorithms**
  - **Long only**
  - **Sign(R)**

$$A_t = \text{sign}\left(r_{t-252:t}\right)$$

  - **MACD signal**

$$A_t = f\left(\text{MACD}_t\right)$$

$$f\left(\text{MACD}\right) = \frac{\text{MACD}\, e^{-MACD^2/4}}{0.89}$$

Experiments

- **Hyperparameters**

Table 1: Values of hyperparameters for different RL algorithms.

Model	α_{critic}	α_{actor}	Optimiser	Batch size	γ	bp	Memory size	τ
DQN	0.0001	-	Adam	64	0.3	0.0020	5000	1000
PG	-	0.0001	Adam	-	0.3	0.0020	-	-
A2C	0.001	0.0001	Adam	128	0.3	0.0020	-	-

Experiments

- **Portfolio**
  - **Simple portfolio by giving equal weights to each contract**

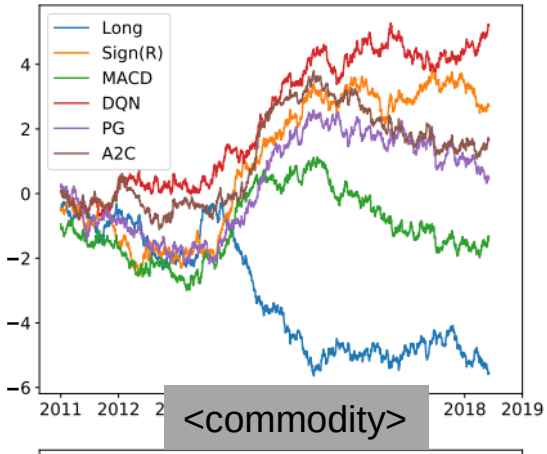
$$R_t^{\text{portfolio}} = \frac{1}{N} \sum_{i=1}^{N} R_t^i$$

- **Metrics**

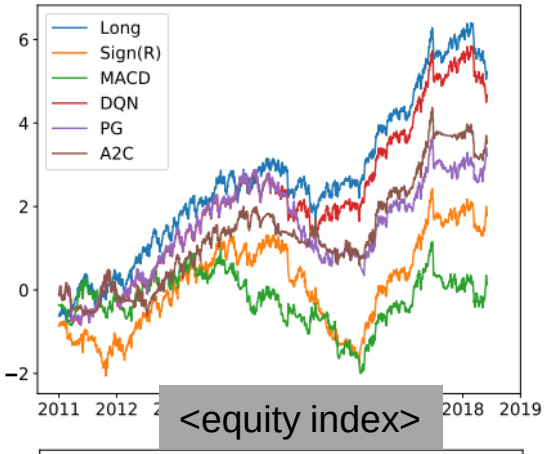
1. $E(R)$: annualised expected trade return,
2. $\text{std}(R)$: annualised standard deviation of trade return,
3. Downside Deviation (DD): annualised standard deviation of trade returns that are negative, also known as downside risk,
4. Sharpe: annualised Sharpe Ratio ($E(R)/\text{std}(R)$),
5. Sortino: a variant of Sharpe Ratio that uses downside deviation as risk measures ($E(R)/$Downside Deviation),
6. MDD: maximum drawdown shows the maximum observed loss from any peak of a portfolio,
7. Calmar: the Calmar ratio compares the expected annual rate of return with maximum drawdown. In general, the higher the ratio is, the better the performance of the portfolio is,
8. % +ve Returns: percentage of positive trade returns,
9. $\frac{\text{Ave. P}}{\text{Ave. L}}$: the ratio between positive and negative trade returns.

Experiments

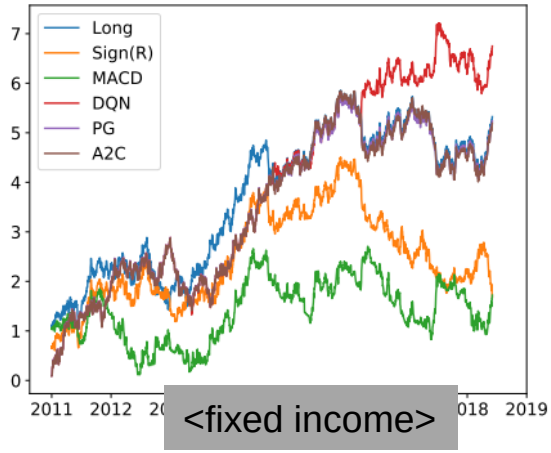
- **Results**



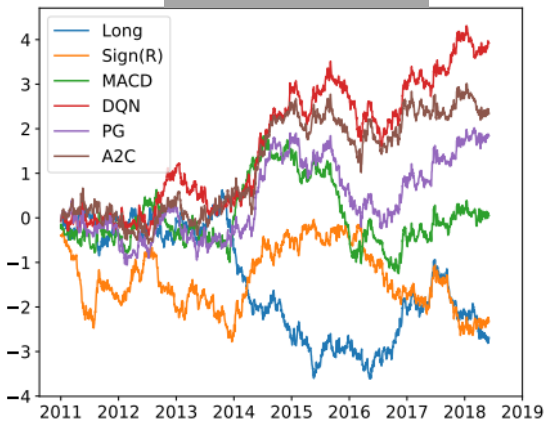
<commodity>



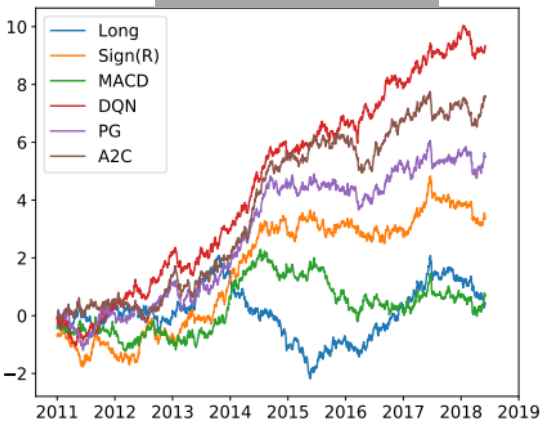
<equity index>



<fixed income>



<FX>



<portfolio>

Experiments

- **Results**

Table 2: Experiment results for the portfolio-level volatility targeting.

	E(R)	Std(R)	DD	Sharpe	Sortino	MDD	Calmar	% of + Ret	Ave. P / Ave. L
					Commdity				
Long	-0.710	0.979	0.604	-0.726	-1.177	0.350	-0.140	0.473	0.989
Sign(R)	0.347	0.980	0.572	0.354	0.606	0.116	0.119	0.494	1.084
MACD	-0.171	0.978	0.584	-0.175	-0.293	0.190	-0.060	0.486	1.026
DQN	**0.703**	0.973	**0.552**	**0.723**	**1.275**	**0.066**	**0.501**	**0.498**	**1.135**
PG	0.062	0.982	0.585	0.063	0.106	0.039	0.023	0.495	1.029
A2C	0.223	0.955	0.559	0.234	0.399	0.141	0.091	0.487	1.093
					Equity Index				
Long	**0.668**	0.970	0.606	**0.688**	**1.102**	0.132	**0.509**	**0.542**	0.948
Sign(R)	0.228	0.966	0.610	0.236	0.374	0.344	0.077	0.528	0.930
MACD	0.016	0.962	0.618	0.017	0.027	0.311	0.006	0.519	0.927
DQN	0.629	0.970	0.606	0.648	1.038	0.161	0.381	0.541	0.944
PG	0.432	0.967	0.605	0.447	0.714	0.242	0.185	0.529	0.960
A2C	0.473	0.929	**0.593**	0.510	0.798	**0.124**	0.328	0.533	**0.962**
					Fixed Income				
Long	0.680	0.975	0.576	0.698	1.180	**0.061**	0.444	0.515	1.054
Sign(R)	0.214	0.972	0.592	0.221	0.363	0.080	0.083	0.504	1.019
MACD	0.219	0.967	0.579	0.228	0.380	0.065	0.123	0.486	**1.101**
DQN	**0.908**	0.972	**0.562**	**0.935**	**1.617**	0.062	**0.543**	0.515	1.098
PG	0.705	0.974	0.576	0.724	1.225	**0.061**	0.436	**0.517**	1.052
A2C	0.699	0.979	0.582	0.714	1.203	0.067	0.408	**0.517**	1.048

Experiments

- **Results**

	FX								
Long	-0.344	0.973	0.583	-0.353	-0.590	0.423	-0.097	0.491	0.979
Sign(R)	-0.297	0.973	0.592	-0.306	-0.502	0.434	-0.111	0.499	0.954
MACD	0.006	0.970	0.582	0.007	0.011	0.329	0.002	0.493	1.029
DQN	**0.528**	0.967	**0.553**	**0.546**	**0.955**	0.183	**0.313**	**0.510**	**1.051**
PG	0.248	0.967	0.566	0.257	0.438	0.240	0.124	0.506	1.021
A2C	0.316	0.963	0.563	0.328	0.561	**0.165**	0.201	0.507	1.026
	All								
Long	0.055	0.975	0.598	0.058	0.092	0.071	0.013	0.520	0.933
Sign(R)	0.429	0.972	0.582	0.441	0.737	0.038	0.201	0.510	1.031
MACD	0.089	0.978	0.582	0.091	0.153	0.008	0.035	0.493	1.043
DQN	**1.258**	0.976	**0.567**	**1.288**	**2.220**	**0.002**	**1.025**	**0.543**	**1.043**
PG	0.740	0.980	0.593	0.754	1.247	0.012	0.480	0.533	0.990
A2C	1.024	0.975	0.573	1.050	1.785	0.007	0.685	0.538	1.021

Overall, DQN obtains the best performance among all models and the second best is the A2C

Experiments

- **Results**

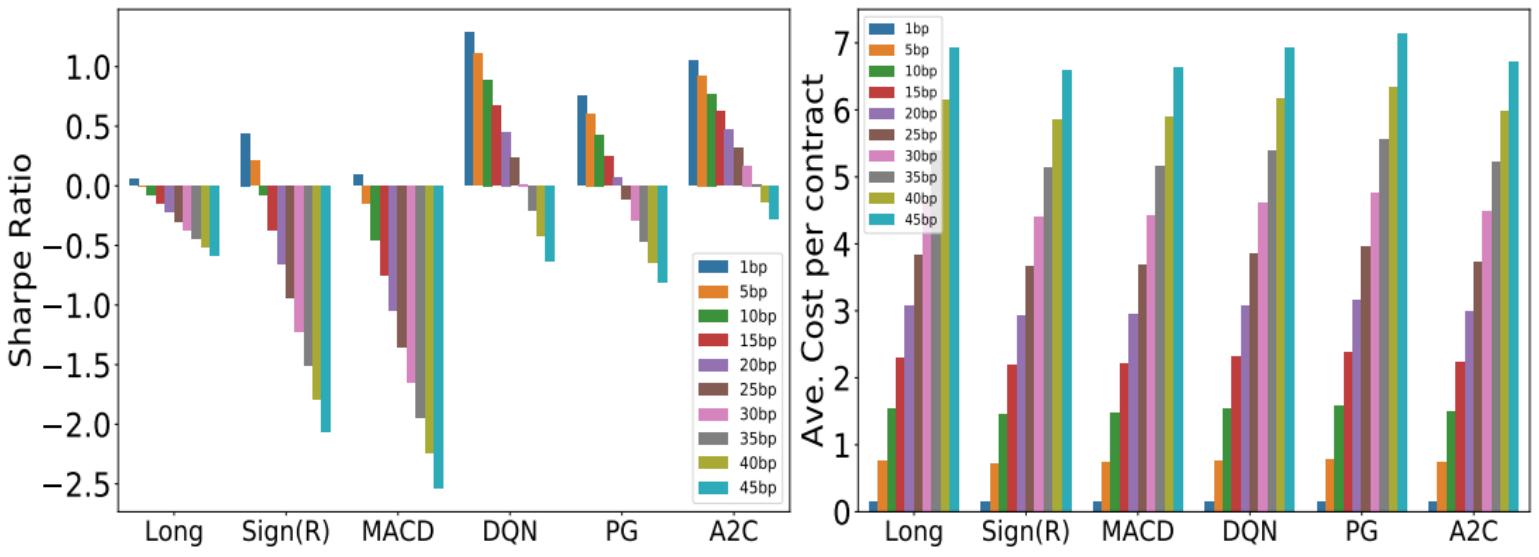


Figure 2: Sharp Ratio (**Left**) and average cost per contract (**Right**) under different cost rates.

Conclusion

- **Adopting RL algorithms to learn trading strategies for continuous futures contracts**

- **Utilizing features from time series momentum and technical indicators to form state representations**

Thank you for Listening