Project Assignment: CCE Aug-Dec 2019 – Reinforcement Learning

Title: Application of Reinforcement learning towards Algorithmic trading and portfolio risk management

Author: Vyasraj Satyanarayana

Submission date: 15th Dec 2019

Summary

Adaption of Reinforcement learning techniques for trading and portfolio management are being extensively researched and studied in the world of Finance. This project assignment attempts to apply Q-learning technique and come up with a model for trading and risk management on financial instruments traded on Indian National stock exchange (NSE).

Algorithm considered here is Temporal-difference learning technique, especially off-policy Q-learning TD-control mentioned in [1]. Initial intent of the project was to develop a hierarchical RL scheme for trading and risk management. Goal of risk management method would be limit the exposure to volatile instruments at any point time, while goal of RL trading system is to enable effective trades maximizing gains. I was able to only develop the trading scheme based on RL and higher Risk layer management is pending further investigation

RL techniques for trading have yielded mixed results, requiring further analysis. Stationarity assumption of the data is also very likely one of the causes for the mixed results.

Next Steps: (Continue this assignment on to next Semester learning of Deep RL course) for

a) Verify whether better convergence can be obtained with Stocastic gradient TD methods
b) Change the Reward/State structure to include instantaneous gains/losses
c) Use Market depth (Buy and Sell Queue data) for improving the results
d) Benchmark the performance, based on Sharpe ratio.

Method and Model

Basic Trade Data

Data of Trades in NSE Instruments is available as a "tick" information. Typically a tick is generated every trade, which involves a settlement between a buyer and a seller of that instrument at a particular price. This results in huge data set for an active exchange like NSE, with volumes of data typically of the order of ~1G ticks/instrument/day.

For this analysis, we consider a summarized "tick" every minute, instead of every trade. This is represented as {Open, High, Low, Close, Volume} of trade data every minute. This results in reduced summarized data set = { 60min*Trading duration = 6hr/day} = ~360 ticks per day/per instrument.

Data set for last 15 years is available from [**Error! Reference source not found.**]

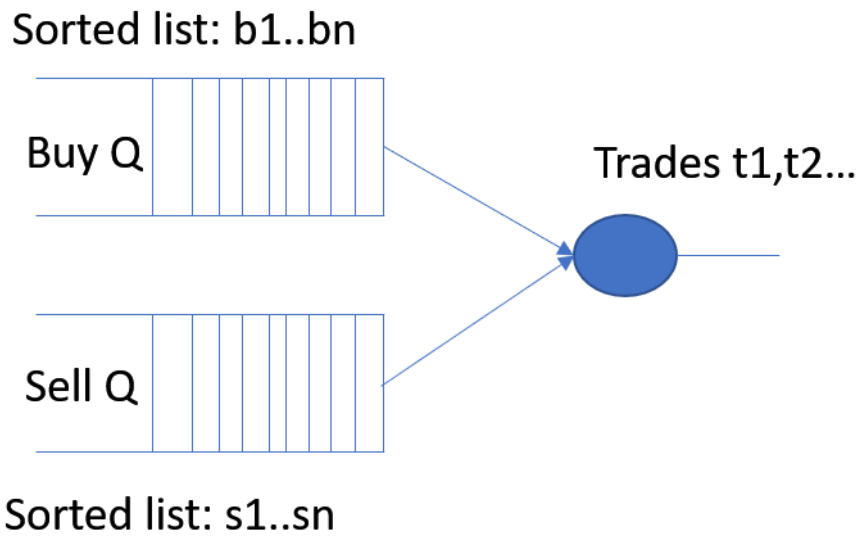


Figure 1Simplified Trade model

A Simplified trade model is captured in Figure 1. Each of the buyer and seller bids are queued in a sorted list. A trade happens when the buyer ask and seller bid matches along with the volume.

Every tick in the Data set captures the summarized {open trade, high trade, low trade, closing trade and volume of trading quantity} for the corresponding minute.

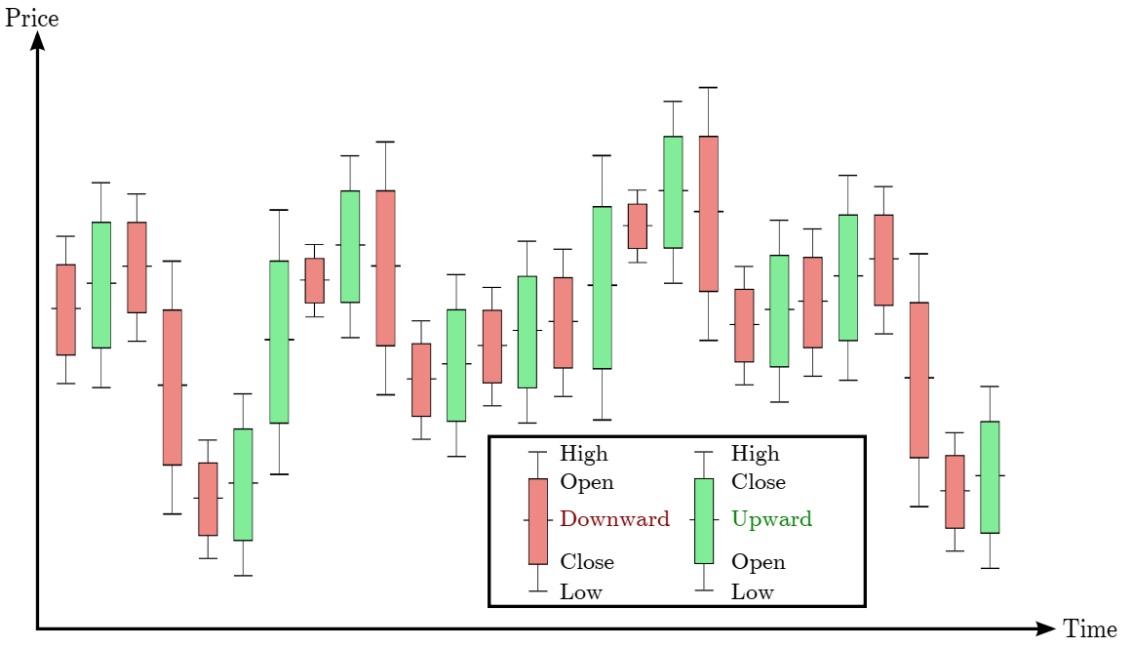


Figure 2OHLC CandleStick of trade graph

RLAgent as participant

RL Agent will be a participant in the market, executing Buy and sell transaction. RL Agent takes a position when it "buys" an instrument at particular price and closes the position when it sells the bought instrument.

Performance metric: Profit and loss (PnL) for a trade pair {Buy/Sell}. Eventual PnL would be a sum of all the trading pair events.

PnL = {Sell Value – Buy Value – slippage}

Slippages constitutes cost involved for each trade. For the simulation we consider 0 slippages. It is also assumed that trades are settled at the close price of the tick, which is not necessarily true, in the real world. This adds to the slippages.

Modelling:

We model a 2-layer hierarchical RL Agent as follows

a) Top level RL Risk management layer
b) Next level of RL agent for instrument trade management

Figure 3 shows the hierarchical layering of RL-Agent.

Goal of the RL agent: Beat and provide a return greater than the Risk-free return.Risk-free return is the theoretical return attributed to an investment that provides a guaranteed return with zero risk. The risk-free rate represents the interest on an investor's money that would be expected from an absolutely risk-free investment over a specified period of time. Investors measure the risk free returns against the return of Government treasury bonds.

The expected return of the portfolio is calculated as a weighted sum of the individual assets' returns. If a portfolio contained four equally-weighted assets with expected returns of 4, 6, 10, and 14%, the portfolio's expected return would be:

(4% x 25%) + (6% x 25%) + (10% x 25%) + (14% x 25%) = 8.5%

Typically volatile instruments carry higher downward risk while maximing gains. Non-volatile instruments yield less gain, but carry less risk as well.

Risk management would classify every instrument as volatile and non-volatile as well as distribute the given trading amount in appropriate ratiosbetween the volatile and non-volatile instruments.

Once the trading instruments have been identified, the RL agent would try to learn the properties of that instrument and aim for a maximum return, in the given constraint.

This project implements only the RL-Agent which trades on instruments. Risk management is pending implementation.

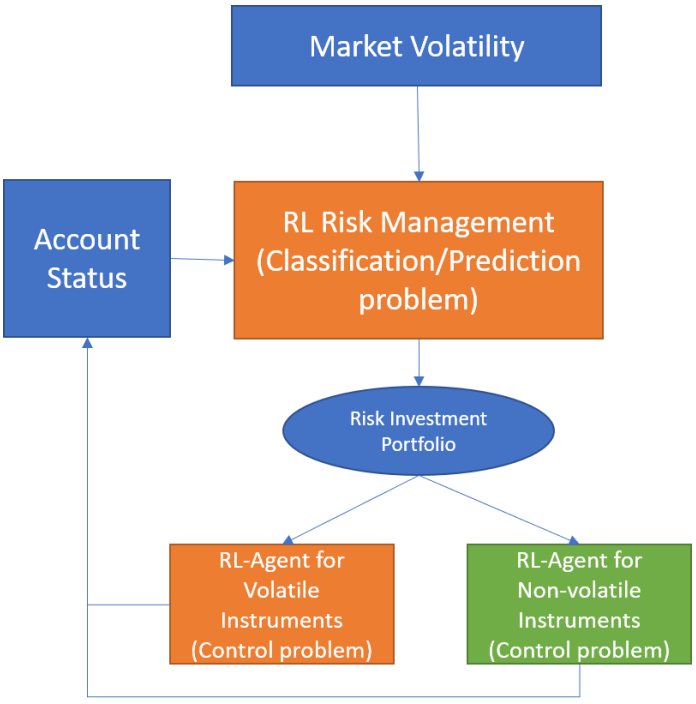


Figure 3Modelling of Hirerchical RL Agents

RL model

State space and Action space:

Table 1State space

States						Instrument position	Comment
Position	**Short term trend {last 15 minutes}**	**Short term trend { last 30 minutes}**	**Previous day trend**	**Previous 2-day trend**	**Previous 3-day trend**		
No position	up down Flat	up down Flat	up down Flat	up down Flat	up down Flat	0	No position. This is the position in the beginning and end of the day, after every episode
Bought	up down Flat	up down Flat	up down Flat	up down Flat	up down Flat	+1	Bought Position only in instrument

State Space is represented as 6-tuple consisting of {Position, $Trend_{15min,}$ $Trend_{30min}$, $Trend_{1day,}$$Trend_{2day}$,$Trend_{3-day}$}. This is shown in Table 1

Trend is computed as following:

T_{diff} = {Current Price – Reference Price}/ReferencePrice*100

Consider a Treshold T_h

$Trend_x$➔UP , if T_{diff}>= T_h

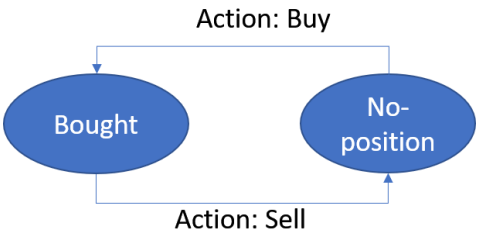
$Trend_x$➔DOWN , if T_{diff}<= T_h

$Trend_x$➔FLAT , if T_{diff}within ($-T_{h,}$ $T_{h)}$)

Position are {Bought or sold}

S_t = {Position, $Trend_{15min,}$ $Trend_{30min}$, $Trend_{1day,}$$Trend_{2day}$,$Trend_{3-day}$}

Actions are {Buy, Sell}



RLAgent-reward Structure

The main goal of the RL agent is to better the Risk free return. Towards the reward structure is defined as follows:

R_{ref} = risk free return
PnL = sell_price – buy_price

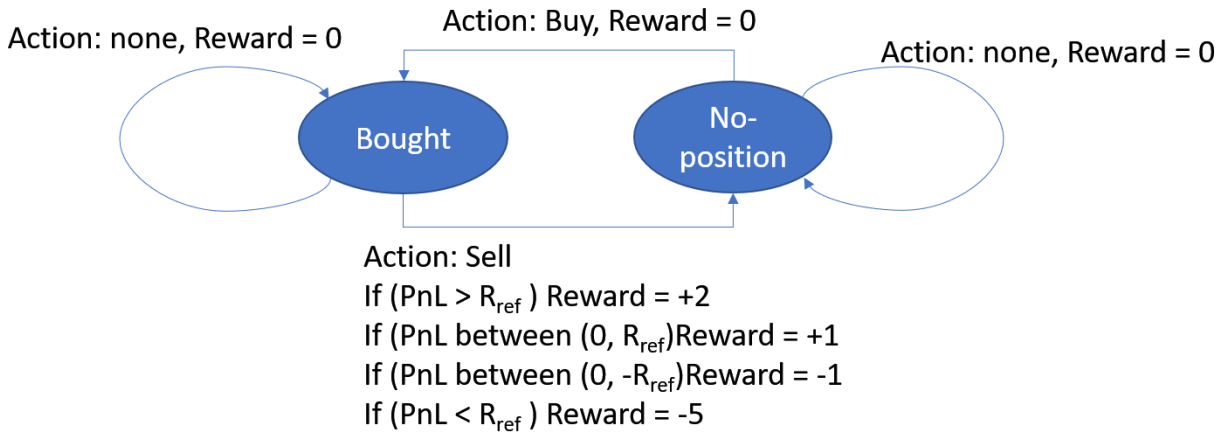


Figure 4Reward Structure

Data-set:

Data-set for an instrument is divided into 2 separate sets. One set as Training Samples and other set for measuring performance of the RL-agent.

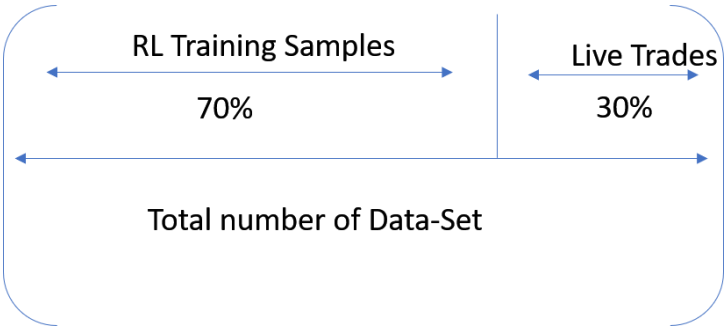


Figure 5Data-set division

Algorithm:

Off policy Q-Learning algorithm based onTD(0) is considered. 70% samples are trained using this algorithm

At End of Day, trades are eventually closed and the risk is not carried over. Every episode begins at 30 minutes into the day (as to compute $Trend_{30min}$, one needs to wait for 30 minutes) and ends by EOD tick.

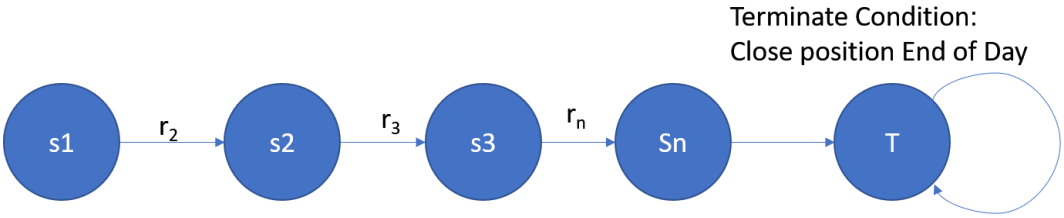


Figure 6 State Transition diagram

Simulation Assumptions:

a) Trades happen at either Open or Close price
b) Slippages are assumed to zero
c) We trade a single unit of instrument. This assumption may not scale for larger volumes

Alpha = 0.1

Discount factor = 0.1

Results

RL Agent for Trading – Non-Volatile instrument

Trends do exist in Trading. If the RL agent can identify such trends and effectively take a position, it is expected to beat the market.

Example of a non-volatile Stock.

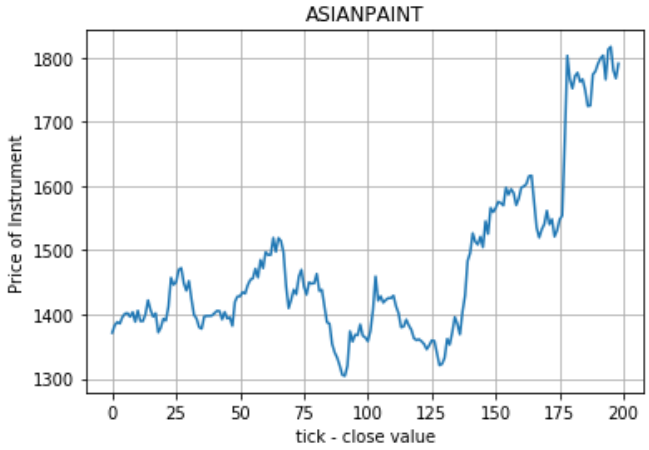


Figure 7 Day-tick of a non-volatile NSE instrument ASIANPAINT

Typical Every minute tick graph of a non-volatile instrument is captured in Figure 8.

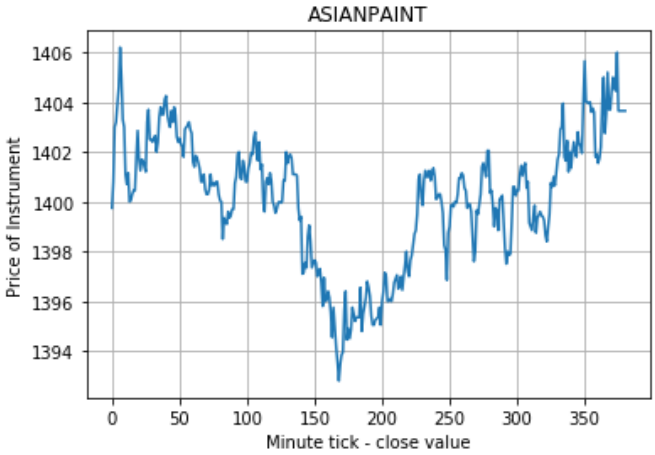


Figure 8 Every minute tick of a typical non-volatile stock

For this graph: { Variance: 6.9, Standard deviation: 2.63, Mean: 1400.13} Also {High – close = 0.86%}

Performance of the RL Agent – Non-Volatile instrument:

Example performance of a RL Agent is captured in

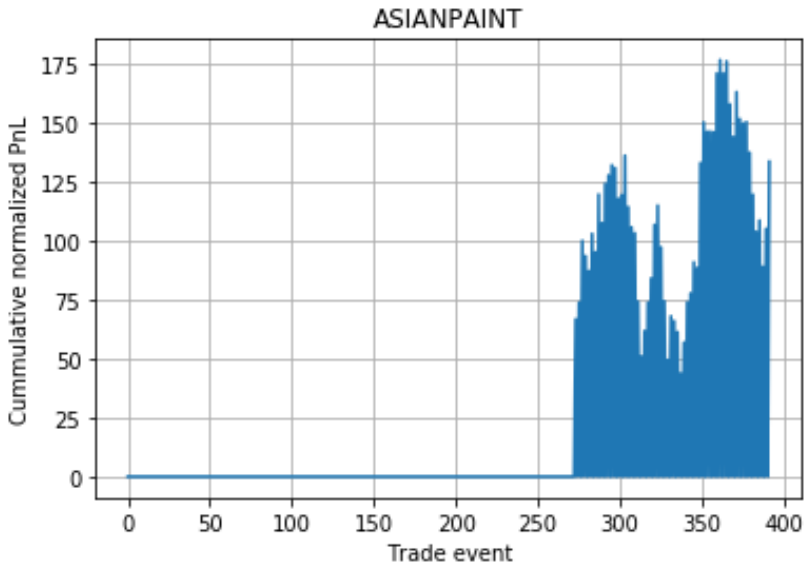


Figure 9CummulativePnL due to position of RLAgent

Approximate returns are ~10% in Figure 9 and the RL agent consistently been above average return value.

RL Agent for Trading – Volatile instrument

Performance of the RL agent for volatile instrument seems questionable.

Consider the instrument show in Figure 10

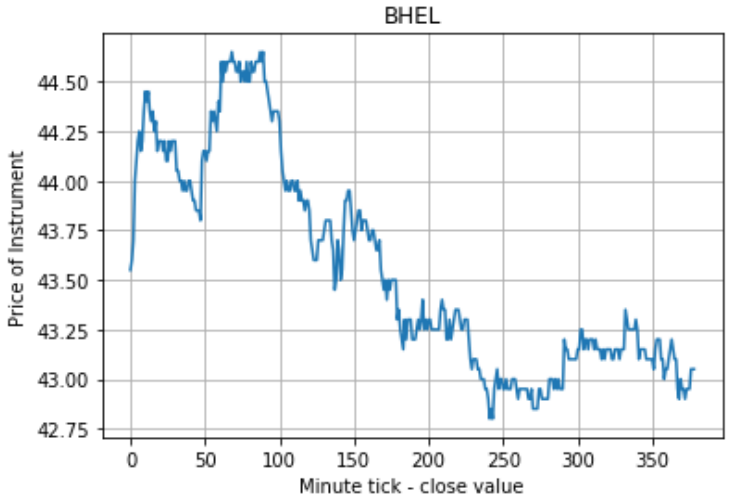


Figure 10Volatile instrument

Variance of this graph: {Variance: 0.28, Standard deviation = 0.53, Mean 43.55} . {High-Close ~ 4%}. When compared to Figure 8.

RL Agent performance is for the above Figure 8 is shown in

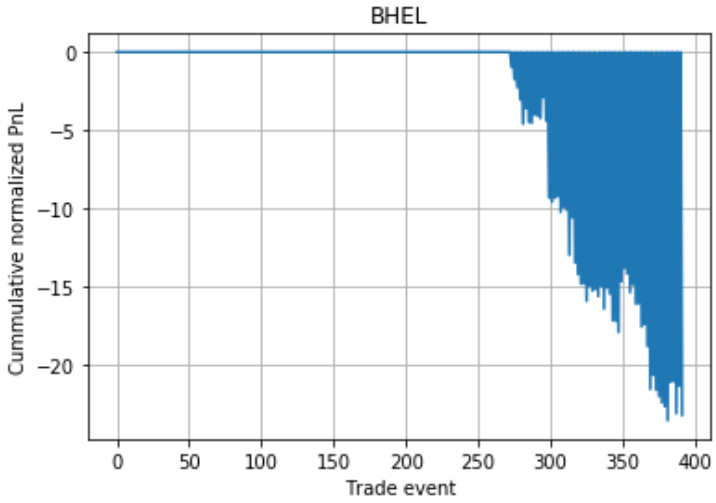


Figure 11RL-Agent performance for a volatile instrument

RL agent performance very badly and is in consistent losses.

Conclusion

It was observed for many non-volatile instruments where the range of {High-Low} is less, RL-agent performed reasonably well. For volatile instruments, assumption of MDP fails and possibly due to non-stationarity of data. A larger "alpha" factor would possibly help the algorithm converge better.

Further experiments on changing the State space based on short term returns could potentially yield better results.

References

1. Sutton, R.S., Barto, A.G. (1998). *Reinforcement Learning: An Introduction*. MIT Press.
2. Fischer, Thomas G., 2018. "Reinforcement learning in financial markets - a survey," FAU Discussion Papers in Economics 12/2018, Friedrich-Alexander University Erlangen-Nuremberg, Institute for Economics.
3. Cumming, J., Alrajeh, D., Dickens, L., 2015. An investigation into the use of reinforcement learning techniques within the algorithmic trading domain. Master's thesis, Imperial College London
4. I. Halperin, "QLBS: Q-Learner in the Black-Scholes (-Merton) Worlds", https://papers.ssrn.com/sol3/papers.cfm?abstract id=3087076 (2017).
5. Yuqin Dai, Chris Wang, Iris Wang, Yilun Xu Stanford University, Reinforcement Learning for FX trading http://stanford.edu/class/msande448/2019/Final_reports/gr2.pdf

Source-code: https://github.com/s-vyasraj/RL-trading