A Deep Reinforcement Learning Approach for Automated Cryptocurrency Trading

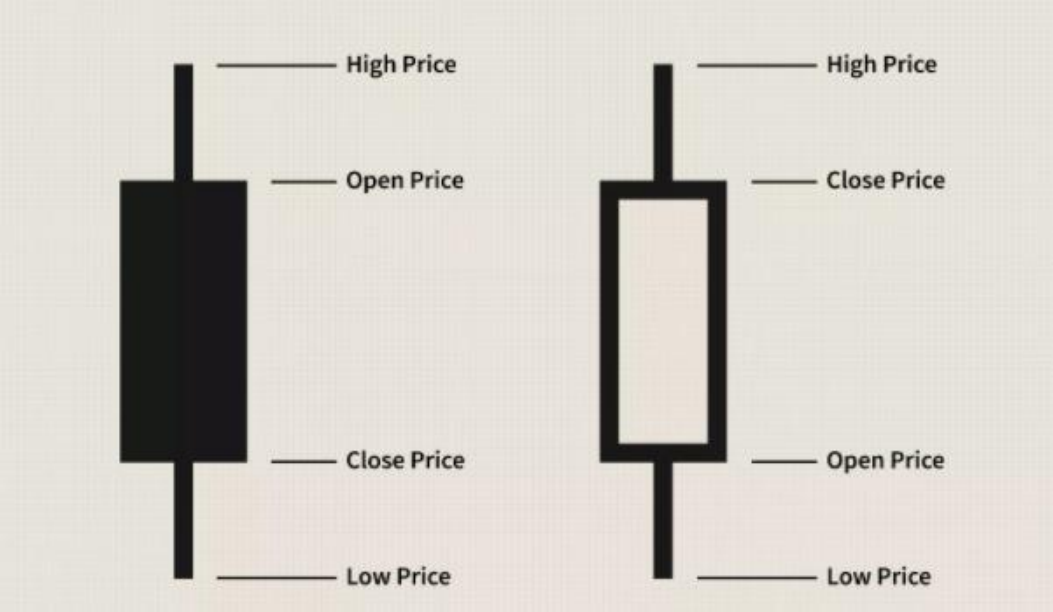
Giorgio Lucarelli and Matteo Borrotti

Universit`a degli Studi di Milano-Bicocca, Milan, Italy

Istituto di Matematica Applicata e Tecnologie Informatiche, IMATI-CNR, Milan, Italy

Cryptocurrency

Candlestick



Candlestick chart

Reinforcement Learning

Different Q-networks are available in literature:

- **Deep Q-Networks** (DQNs) - DQNs stabilize the training of action value function approximation with deep neural networks, in particular Convolutionary Neural Networks (CNNs), using experience replay and target network.
- **Double Deep Q-Networks** (D-DQNs) **-** D-DQN improved DQN avoiding over-estimation. In D-DQN a greedy policy is evaluated in accordance with an online network and a target network is used to estimate its value.
- **Dueling Double Deep Q-Networks** (DD-DQNs) - is based on a dueling network architecture to estimate value function $V(s)$ and the associated advantage function $A(s, a) = Q(s, a) - V(s)$, and then combine them in order to estimate $Q(s, a)$. In DD-DQN, a CNN layer is followed by two streams of fully connected (FC) layers, used to estimate the value function and the advantage function separately; then the two streams are combined to estimate the action value function.

Q-learning Trading System

The proposed Q-learning trading system is based on **Double Deep Q-Networks** and **Dueling Double Deep Q-Networks.**

The Q-learning trading system rewards the agent with two possible functions - Sharpe ratio and simple Profit function.

Sharpe ratio

$$\begin{cases} s_{p_t} \geq 4 \rightarrow reward = +10 \\ 1 < s_{p_t} < 4 \rightarrow reward = +4 \\ 0 < s_{p_t} \leq 1 \rightarrow reward = +1 \\ s_{p_t} = 0 \rightarrow reward = 0 \\ 0 < s_{p_t} \leq -1 \rightarrow reward = -1 \\ -1 < s_{p_t} < -4 \rightarrow reward = -4 \\ s_{p_t} \leq -4 \rightarrow reward = -10 \end{cases}$$

Profit function

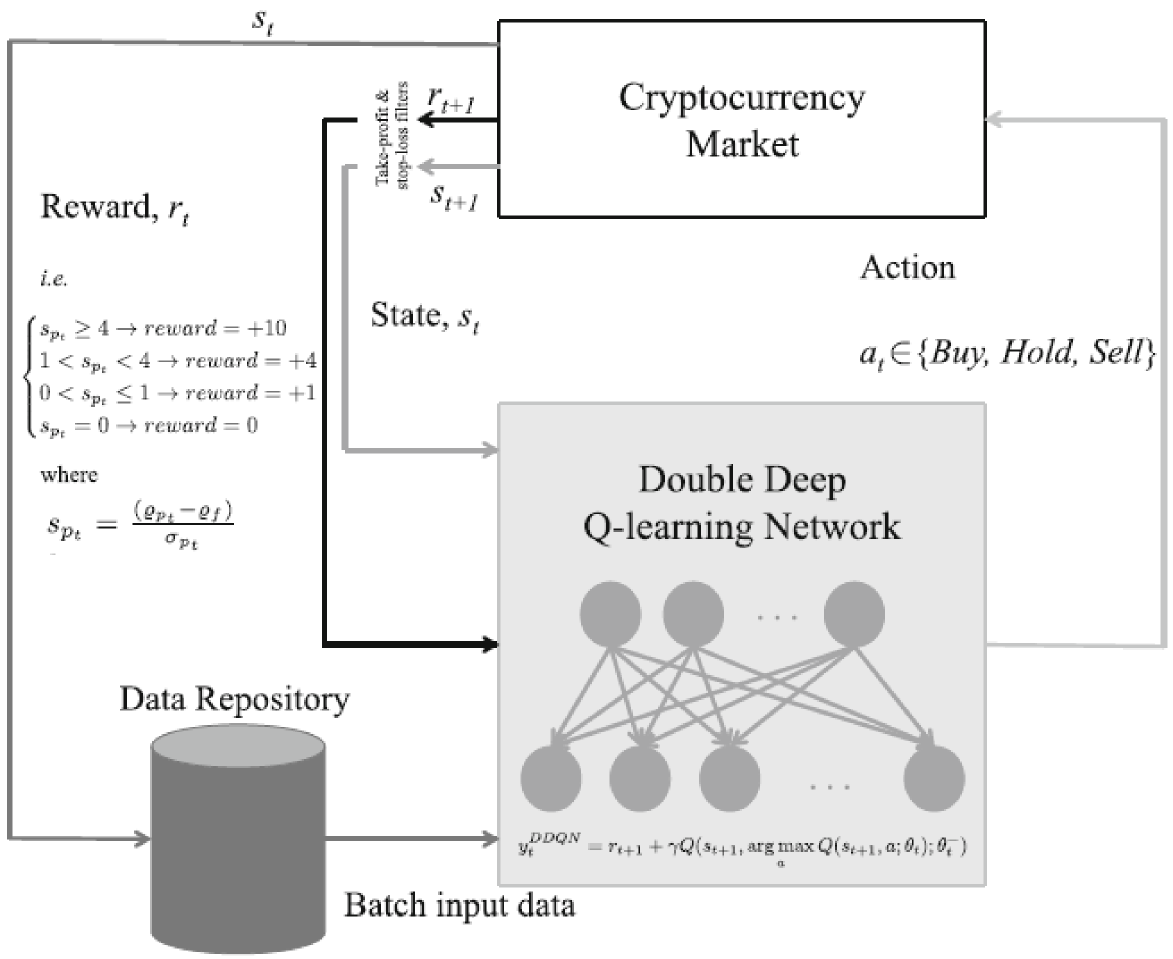
$$\begin{cases} g_{profit} > 0 \rightarrow reward = 1 \\ g_{profit} = 0 \rightarrow reward = 0 \\ g_{profit} < 0 \rightarrow reward = -1 \end{cases}$$

Experimental Data and Results

Average performance over the 10 trading periods.

Trading system	Avg. return (%)	Max. return (%)	Min. return (%)	St. dev.
*Profit*D-DQN	3.74	21.31	−10.74	4.87
*Profit*DD-DQN	4.85	17.34	−8.49	5.10
*Profit*DQN	2.32	22.59	−17.97	7.93
*Sharpe*D-DQN	5.81	26.14	−5.64	5.26
*Sharpe*DD-DQN	3.04	13.03	−8.49	3.81
*Sharpe*DQN	1.83	15.80	−9.29	5.46

Double Deep Q-learning System with Sharp reward function



Conclusions and Future Work

- Positive return
- Limitations
- To study the impact of social media and incorporating news and public opinion into the Deep Reinforcement Learning approach
- Secure and trustworthy anomaly detection system