Accurate Price Prediction by Double Deep Q-Network

Mohammad-Reza Feizi-Derakhshi [1], Bahram Lotfimanesh [2], Omid Amani [3]

[1] Computerized Intelligence Systems Laboratory, Department of Computer Engineering, University of Tabriz, Iran.
[2] Computerized Intelligence Systems Laboratory, Department of Computer Engineering, University of Tabriz, Iran.
[3] Department of Computer Engineering, University College of Nabi Akram, Tabriz, Iran.

[1] mfeizi@tabrizu.ac.ir
[2] blotfimanesh@gmail.com
[3] omidamany1366nnn@gmail.com

Abstract For more than several decades, time series data have been in the center of attention for scholars to predict the future prices of the markets, the most fundamental and challenging of which has been the prediction of the price of the stock market. It is of great importance to note that the algorithms with the fewest errors in price predictions are more applicable. There have been more methods suggested for price prediction in the stock markets: time series data analysis, mathematical and statistical analysis, signal processing, pattern recognition and machine learning. One of the demerits of the aforementioned methods is failing to recognize sudden change of prices, in this regard, experiencing more errors is the consequence of such demerit. In this regard, to have the error solved, the DDQN algorithm, consisting of deep neural networks which includes LSTM-CNN layers, has been employed. Confronting price fluctuations, the agent has the privilege of having better performance by employing the advantages of LSTM-CNN layers. In this research, the algorithm has been carried out over Iranian Gold Market, including six various types of Gold, from 2009 to 2020. The results reveal the point that the given method is more precise in comparison with other suggested methods confronting sudden changes in prices.

Keywords: Stock, Stock Price, Prediction, Neural Network, Deep Reinforcement Learning, DDQN, LSTM, CNN.

1 Introduction

Recently, machine learning has come to the fore in different aspects, one of the most important of which is stock marketing and stock trading. The exact prediction of stock price is of great challenge for analysts and financial investors [1]. The prediction is bound with political circumstances, global economic condition, financial reports, and the financial turnover of the companies [16]. The oscillation in stock market price has undergone some dramatic shifts because of global changes and the manner of investors [4]. In this regard, for maximizing the profit and keeping the loss in minimum rate, some technical approaches for the prediction of price value of the stock market, considering recent annual trends, can be so beneficial to have the stock market changes under scrutiny [8]. In comparison to the ancient approaches, learning and making decisions in various periodical time spans are of the fundamental challenges in stock markets in that machine learning techniques have ameliorated the efficiency of predicting process. (60 to 86 %) [9]. To confront such challenges, it can be mentioned that reinforcement learning as a sub-branch of machine learning has been presented to alleviate such challenges. The approach has been inspired by biological theory coined by Schultz, W., Dayan, P., & Montague, P. R in 1997 [12]. Deep reinforcement learning, on the other hand, has merged deep learning by reinforcement learning resulting in simulation of perception of humans and learning methodologies [10]. Deep neural network has some pros and cons, as for the former, having strong perception and feature extraction ability can be pointed out, while for the
latter, being indecisive for decision-making processes can be highlighted. To have eliminated such deficiency, reinforcement learning has come to the fore to cross out such weak points in deep-learning process [16]. Mnih in one of the protagonists who suggested a deep Q network, a network of convolutional neural network in reinforcement learning [11]. Double Q network was coined by Hasselt leading to providing buffer in environment[5]. Following this vein, Double DQN was provided which resulted in two sub-models within to ameliorate agent efficiency in environment [15]. In this study, a special deep reinforcement learning has been presented.

It is an algorithm in which the combination of LSTM and CNN neural network have been employed for the prediction of gold price, which has not been highlighted by scholars before. To hit this target, a given agent employs a deep neural network which includes LSTM-CNN layers. The merits of the LSTMCNN neural network layers are learning long-term dependency records and various feature extractions for precise and concise prediction of the agent. In this regard, for boosting learning process, various windowing has been employed to broaden agent’s vision. In this respect, the given agent will suffice by appropriate amount of information for learning and providing new states [7].

Briefly, the most highlighted implication of this article can be summarized as:

- Pioneering analysis and performance of Double deep Q network including LSTM-CNN neural network layers which has been case studied over the predication of Gold price.
- Analyzing the proposed model in this study by analogy of other models has paved the way for better efficiency of the predictions.
- Learning and amelioration of the agent’s efficiency by considering provided information and features and following up the given learning process to hit the assigned target of the agent. In other terms, the capricious shocking stock market oscillations will be tackled much easily by the given participant.

In the following, for the sake of brevity, some scientific approaches will be under scrutiny as a literature review and the proposed model will be discussed following the general implications of the study, in the last section, the further studies and final results will be pointed out.

2 Review of the Literature

In the recent studies, it can be easily noticed that deep reinforcement learning plays a significant role for the prediction of stock market price, however, there exists other algorithms which can be in the center of attention noticeably. In this study, one of these algorithms will be discussed. It is the Deep Reinforcement Learning consisting of LSTM-CNN neural network layers. Stock market studies reveals the facts that knowledge and being familiar with the price of an item in the stock market in long-term periods and saving up precious and influential information can result in solid and concise price prediction [8]. Providing appropriate circumstances for the agent, considering various and long-term perspective based on training data, which is one of the fundamental features of LSTM has resulted in soaring up the accuracy in price prediction. In this section, we will have a brief look at the pervious review of the literature to have a better understanding over the proposed model in the study.

Z. He.et all [6] have proposed a supervised model which includes LSTM-CCN layers for gold price prediction. In this model, the layering process, the number of layers, and other dependent variables have been scrutinized, and the data set of WGC1 has been employed that consists of 10471 records from 1978.12.29 to 2019.04.15. Firstly, the input of neural network is normalized by SCIKLITL learn [2] and then the windowing process is applied for training and presented for the model. It is the best model of the all presented models in this section consisting of two LSTM layers, one attention layer, one CONV2D layer, one MAX POOLING, and the ultimate fully connected layer. The evaluation criteria is based on RMAE, MAPE, and RMSE. In this regard, in Z. He approach, deep reinforcement learning has been ignored totally. Salvator.et all [3] have been suggested a method by which the given agents have been trained in various epochs in the environment, and the agent who has hit the maximum grade is nominated for further actions. The aforementioned reason is the strongest aspect of Salvatore Carta.et all [3] approach. It is wise to note that in his model, fully connected neural network layer is employed and two specific data sets of DAX and S&P500 have, indeed, been used for model analysis. Li, Y.et all [10] have offered a model in which the analysis and comparisons of deep reinforcement learning algorithms have been studied. Based on the gained profits, the algorithms have been evaluated. The employed data set has been extracted from KAGGLE dividing in train and test. Each data set is analyzed by deep Qnetwork (DQN), double deep Q network (DDQN), and dueling double deep Q network (dueling DDQN). The neural network employed in the above-mentioned models is fully connected.

S. Selvin.et all [13] have suggested and probed three algorithms for stock price predictions. In this approach, the evaluation of algorithms based on their faults have been analyzed. The

1www.gold.org
whole data set consist of the exact stock market price for 1721 stock market NSE companies from July 2014 to June 2015. In this article, the sliding window approach has been carried out for a 100-minute sized window and 90-minute overlapped window and the prediction for 10 minutes ahead. The size of windowing has been identified based on trial and effort style. Having considered the results, it can be noticed that CNN model has had much better efficiency compared to other models. The given results in the proposed model of this study reveals the point that DDQN model with the LSTM-CNN neural networks layer has boosted the accuracy of prediction and has plunged the errors.

3 Our Approach

There has been much attention over trading in stock market and a precise and dynamic system to minimize the risks in trading markets. Machine learning, specifically deep reinforcement learning with dynamic and complete algorithms, has a massive potential for attention and development to run with time series data. The aforementioned algorithm has been ignored so in this study, a model has been suggested to merge reinforcement learning by neural networks to project the system more effectively. Deep reinforcement learning is an approach by which a given agent is addressed to receive a better perception and have a better efficiency[14]. In this light, boosting the perception of the agent will lead in having a concise agent with minimized error coefficient. The proposed neural network which is used in deep reinforcement learning consists of LSTM-CNN. The objectives of this model are; highly-effective learning of agent through receiving and saving appropriate and related information of stock price, encouraging and punishing of the agent to have an appropriate and concise prediction in the stock market, and tackling unpredicted shocking conditions in the markets. The proposed model is applied to predict the updated daily price of the gold. At the first place, the price of the gold has been obtained online through Iranian Gold Stock website, then the data has been charted in Excel. Having the data loaded in the algorithm and converting it to the array, we have preprocessed on data which have been categorized based on the type of gold. Data has six distinct datasets; Ounce, Gold price, 18 Carat gold, Imami Coin, quarter Coin, and 1g Gold coin. Data set period belongs to 2013.07.24 - 2020.07.02. It is wise to highlight that each type of the dataset consists of 26334 records. In the proposed preprocessed model 1, the input data in Excel is evaluated and preprocessed in that input data are significant to neural networks so that first, based on the Sklearn function, the data is normalized in range (−1,1), then the data is windowed by various size of windowing, next applying Kfold preprocessed in that input data are significant to neural networks so that first, based on the Sklearn function, the data is normalized in range (−1,1), then the data is windowed by various size of windowing, next applying Kfold [2], Train and Test data are prepared to be presented in neural networks.

The categorized data by Kfold is presented to neural network model 2. First, the data is occupied in LSTM that includes 200 neuron and Relu activation. In the second layer, the repeatvector, based on features of output, changes the dimension of data, the output of which is delivered to second layer of LSTM. It is in the second layer containing 200 neurons and Relu activation. Following this respect, the output is presented to Conv1D for extracting features, and then for minimizing the dimension and selecting the best, MaxPooling1D is employed and at the end step of this model, a flatten layer and two dense layer is observed. The output of Neural Network consists of Open price, Close Price, Maximum price and Minimum price of the following day.

Other type of models and approaches fail to have all the conditions of markets observed in the training, and the external factors do not act in vacuum in such models that result in plunging the accuracy of the model, therefore, for tackling the problems, a more effective approach, without the flaws mentioned, is needed. In this respect, we are to employ effective merge of machine learning to detect important features. Given algorithm, ignored so far by scholars, has been resulted in having dynamic, flexible, and precise system. The results reveal the more effective functionality of this model. The proposed model in this study is managed to function more effectively because of having the agent adapt with external factors of stock market price and shocking immediate changes in the market price by recalling fundamental and special information and extracting unique features.

At this level, an ideal model is in the hand which includes appropriate windowed and folded dataset. The next step is training agent based on best available parameters. The proposed algorithm is twofold model; Target Model and the Evaluate Model. In the following, more detailed information will be elaborated based on 3. It is wise to note that for training an agent, a label in the environment is required. To generate the label, we use a complement model.

For achieving such a target, first, we read the record from dataset, and then we exchange them in the format of state to have them transfer to the neural network; in other words, we read the price of three days including: open, maximum, minimum, and close price along with agent reward in each iteration to send them to neural network, consequently, in this algorithm, we need to define the size of read-data called Batch-size, the size of which is presented as 64 by default. It implies the fact that learning happens if and only if there are 64 records in the buffer, therefore, in each training step, we read records as big as the Batch-size capacity from buffer and we transfer them to neural network to predict some values and based on these values, training is processed. At this level, some

\footnote{www.gold.org}
values are predicted for next-state by Target Model, and error of these values is calculated by Loss-Function Eq. 6. To do so, we need to define some steps to achieve the objectives of this study. The sum of given rewards from each episode has been collected by Return Function which is presented in Eq. 1:

\[ G_t = R_{t+1} + \gamma R_{t+2} + \gamma^2 R_{t+3} + \cdots = \sum_{k=0}^{\infty} \gamma^k R_{t+k+1} \]

For each Episode: S1, A1, R2, S2, A2, R3, ...

\[ R_t \]

The amount of reward for agent in T level Gt: the sum of reward for the agent for accomplishing the Action (a), State (s), until the end of an episode γ: 0 ≤ γ ≤ 1; the reason of employing γ is decreasing the effect of future awards in a return value. The return value will be divergent if the episodes do not have an end. To have them converge, we multiply return value by Gama. Obtaining the sum of rewards and states, we can get to the state value Eq. 2:

\[ V_\pi(s) = E_\pi [G_t | S_t] \]  

(2)

![Figure 1: The preprocess procedure. The input data is loaded from dataset, normalized, windowed by various size, and Kfold is applied.](image)

To reach a new state, we have to do an action to get to the Action value Eq. 3:

\[ Q_\pi(s, a) = E_\pi [G_t|S_t = s, a_t = a] \]  

(3)

To hit the optimal value, we, indeed, need to obtain the best value for the states via a definite policy. The policy which has been reached by optimal value is called optimal policy. We need to highlight the point that searching all the policy for finding optimal policy is impossible. To have this problem solved, the Bellman equation is applied as an assist Eq. 4:

\[ V^*(s) = \max_a V_\pi(s) \]  

(4)

Bellman equation was coined by a mathematician Richard Ernest Bellman in 1953. It consists of dynamic programming which is used for decision making process at a given point by overviewing of previous state-value. This model is used for calculating function value in dynamic programming or environment which results in modern reinforcement learning. The most fundamental factors in Bellman equation are:

a. The Action done by the agent called "a"

b. The state provided by action called "s"

c. Reward and punishment provided by any good/bad actions called "R"

d. γ factor; 0 ≤ γ ≤ 1

1www.gold.org
Bellman equation can be written as Eq. 5:

\[
V(s) = E[G_t | S_t = s] = E[R_{t+1} + \gamma R_{t+2} + \gamma^2 R_{t+3} + ... + \gamma R_T | S_t = s]
\]

In the algorithm related to prediction of gold price for training model, we need a label which can be calculated by Eq. 6:

\[
L(s) = \text{Error} \sim (\text{target}(s) - Q_{predict}(s))^2
Q_{target} = R + \gamma L(s)
\]  

(6)

Evaluate model is trained by new labels and tries to update its weights in way to diverge predict values to . The most important factors for agent decision-making process depend on the training style,
Figure 3: Proposed DDQN Model.

Table 1: Parameter Setting for Each Dataset.

<table>
<thead>
<tr>
<th>Dataset</th>
<th>Windowing</th>
<th>Kfold</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ounce</td>
<td>3-1</td>
<td>7</td>
</tr>
<tr>
<td>Gold price</td>
<td>3-1</td>
<td>5</td>
</tr>
<tr>
<td>18 Carat gold</td>
<td>5-1</td>
<td>10</td>
</tr>
<tr>
<td>Imami Coin</td>
<td>5-1</td>
<td>20</td>
</tr>
<tr>
<td>Quarter Coin</td>
<td>5-1</td>
<td>7</td>
</tr>
<tr>
<td>1g Gold coin</td>
<td>7-1</td>
<td>7</td>
</tr>
</tbody>
</table>

and maximizing agent’s reward in the environment. To do so, we training the agent several times in the environment for maximizing the agent’s reward and aging enough experiences from effective parameters. Based on Fig. 3, in each episode including agent’s training, all of the rewards and punishments in each state are stored in a variable called Maxreward, and after finishing each episode, they are sent to Reward Function. If the obtained value is more than the previous stage, training will be repeated again, otherwise, the values will be stored and the training process is done which will lead to testing level.

4 Result and Discussion

After having dataset processed, we go through the parametric examination of each dataset. It has the implication of being certain about the size of windowing and the procedure of division based on Kfold for boosting the accuracy of algorithm for better prediction. In table 1, after some experiments, we have achieved the most effective features of data which are more influential in prediction process. Notably, in all of the experiments, the output is presented as 1 by default.

In this section, the proposed model in this study, DDQN with LSTM-CNN neural network, is probed by two models below:

1. DDQN with CNN-DENSE neural network

2. Supervised model with LSTM-CNN layer

In this case, we noticed that there are various factors contributing to the maximizing our agent accuracy summarized as:

1www.gold.org
1. Maximum level of training to achieve the maximum reward, to do so, the given agent will follow up learning and error correction as much as it is required.

2. Applying LSTM neural network leads into the increase of agent’s vision over price by using influential data in the process of training and this contribution has much more effective impact through shocking changes in price in comparison with other models.

It is obvious that the proposed algorithm has had a minimum error in all of the datasets in the testing level. The results are obtained due to suitable applying Loss Function and optimizing Reward Function and the eye-catching effect of LSTM layer. In the following Fig. 4, the output of the model is presented. All of them has been carried out over Ounce Dataset. Compatibility of label and predict value in Fig. 4a, related to our model, is more than the other models.

Then we applied this model on another dataset, presented in Table 2. In this experiment, the windowing size and the amount of folding have been set based on previous parametric tuning results.

We have employed Histogram to achieve the distribution of numerical label and prediction values. We have subtracted predicted value from label value and sent to the bar charts. Checking the bar charts discloses the fact that our proposed model has had much better efficiency in that experiencing sudden shocking changes have been less than other models. The fact is easily noticeable by bar chart ranges.

Table 2: RMSE-test results on each Dataset. In this case for all models windowing and Kfold set with previous parametric tuning results from table 1.

<table>
<thead>
<tr>
<th>Dataset</th>
<th>Propose Model</th>
<th>DDQN with CNN-Dense</th>
<th>LSTM-CNN</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ounce</td>
<td>0.011</td>
<td>0.013</td>
<td>0.021</td>
</tr>
<tr>
<td>Gold price</td>
<td>0.012</td>
<td>0.016</td>
<td>0.016</td>
</tr>
<tr>
<td>18 Carat gold</td>
<td>0.013</td>
<td>0.014</td>
<td>0.014</td>
</tr>
<tr>
<td>Imami Coin</td>
<td>0.013</td>
<td>0.014</td>
<td>0.015</td>
</tr>
<tr>
<td>Quarter Coin</td>
<td>0.019</td>
<td>0.010</td>
<td>0.012</td>
</tr>
<tr>
<td>1g Gold coin</td>
<td>0.015</td>
<td>0.017</td>
<td>0.059</td>
</tr>
</tbody>
</table>

1www.gold.org
Figure 5: Histogram of the difference between the numerical label and prediction values.

Table 3: Presentation of RMSE decline between proposed and other models.

<table>
<thead>
<tr>
<th>Dataset</th>
<th>DDQN with CNN-Dense</th>
<th>LSTM-CNN</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ounce</td>
<td>%15 ↓</td>
<td>%86 ↓</td>
</tr>
<tr>
<td>Gold price</td>
<td>%30 ↓</td>
<td>%30 ↓</td>
</tr>
<tr>
<td>18 Carat gold</td>
<td>%0.06 ↓</td>
<td>%0.05 ↓</td>
</tr>
<tr>
<td>Imami Coin</td>
<td>%0.06 ↓</td>
<td>%17 ↓</td>
</tr>
<tr>
<td>Quarter Coin</td>
<td>%0.01 ↓</td>
<td>%14 ↓</td>
</tr>
<tr>
<td>1g Gold Coin</td>
<td>%13 ↓</td>
<td>%291 ↓</td>
</tr>
</tbody>
</table>

The most fundamental point in the proposed algorithm is the constancy in all of the predictions, in a way that in all of the datasets predictions have been by far better compared to other algorithms. Another advantage that can be mentioned in this regard is the applying LSTM in deep reinforcement learning neural network in that the application has the ability of considering previous vital parameters to recognize and analyze the most updated daily prices. In Figure 5, we can notice how flexible the algorithm is dealing with the immediate shocking changes, and how adaptable it is matching label and predict price. In figure 5a, LSTM layer is used in DDQN neural network which reveals the fact that the distribution of difference between label and predicted price is up close to 0. In our proposed model the range of distribution is about -60 to 80 compared with other models (5b) which is the range of -60 to 100 and the figure 5c which is in the range of -75 to 125. In the Table 3, the error rate has been decreased in all of the dataset. For instance, in Ounce dataset, the RMSE amount has been plummeted 15% in DDQN with CNN-Dense neural network layer and 86% in LSTM-CNN model.

5 Conclusion

There exist many algorithms in the prediction of stock market price and they have had noticeable results. However, there has been less attention over deep reinforcement learning specially DDQN. We have tried to apply different experiments by the proposed algorithm. We have achieved noticeable results combining LSTM and CNN.
in DDQN neural network. (Discussed in section 4). It is easily observed that perdition in sudden shocking price changes and external factors in the price of gold is done by much better accuracy so that the investors can trade more confidently. Other advantages of the proposed model are much better function of the algorithm confronting price oscillation, having better prediction accuracy, and dealing with sudden shocking price changes compared to other algorithms. We have achieved to this point that the accuracy obtained by this algorithm can be applied to a dynamic system for trading and prediction. To extend such a system will be of great help for all stock market activities which will be discussed in our future researches.

References


1www.gold.org


¹www.gold.org