Stock Price Prediction Using Reinforcement Learning

Chapter · January 2022

DOI: 10.1007/978-981-16-2597-8_6

6 authors, including:

Poonam Rani
Netaji Subhas University of Technology

Jyoti Shokeen
Indira Gandhi Delhi Technical University for Women

Metadata of the chapter that will be visualized in SpringerLink

Book Title	International Conference on Innovative Computing and Communications
Series Title	
Chapter Title	Stock Price Prediction Using Reinforcement Learning
Copyright Year	2021
Copyright HolderName	The Author(s), under exclusive license to Springer Nature Singapore Pte Ltd.

Corresponding Author	Family Name	**Rani**
	Particle	
	Given Name	**Poonam**
	Prefix	
	Suffix	
	Role	
	Division	Department of Computer Engineering
	Organization	Netaji Subhas University of Technology
	Address	Dwarka, New Delhi, India
	Email	poonamrani2017.nsit@gmail.com
Author	Family Name	**Shokeen**
	Particle	
	Given Name	**Jyoti**
	Prefix	
	Suffix	
	Role	
	Division	Department of Computer Science and Engineering UIET
	Organization	Maharshi Dayanand University, Rohtak
	Address	Haryana, India
	Email	jyotishokeen.rs.uiet@mdurohtak.ac.in
Author	Family Name	**Singh**
	Particle	
	Given Name	**Anshul**
	Prefix	
	Suffix	
	Role	
	Division	Department of Computer Engineering
	Organization	Netaji Subhas University of Technology
	Address	Dwarka, New Delhi, India
	Email	anshuls1.co.17@nsit.net.in
Author	Family Name	**Singh**
	Particle	
	Given Name	**Anmol**
	Prefix	
	Suffix	

	Role	
	Division	Department of Computer Engineering
	Organization	Netaji Subhas University of Technology
	Address	Dwarka, New Delhi, India
	Email	anmols.co.17@nsit.net.in
Author	Family Name	**Kumar**
	Particle	
	Given Name	**Sharlin**
	Prefix	
	Suffix	
	Role	
	Division	Department of Computer Engineering
	Organization	Netaji Subhas University of Technology
	Address	Dwarka, New Delhi, India
	Email	sharlink.co.17@nsit.net.in
Author	Family Name	**Raghuvanshi**
	Particle	
	Given Name	**Naman**
	Prefix	
	Suffix	
	Role	
	Division	Department of Computer Engineering
	Organization	Netaji Subhas University of Technology
	Address	Dwarka, New Delhi, India
	Email	namanr.co.17@nsit.net.in

Abstract	With the availability of new data sources and advancement in marketing and financial instruments, stock market returns are a major research area. Stocks have a huge influence on today's economy. A better predictive model is extremely important in stock prediction. The aim of this paper is to investigate the positive effect of reinforcement learning on stock price prediction techniques. Q-learning has been shown to be incredibly effective in various segments, such as cloud scheduling and game automation. This paper demonstrates how the Q-learning technique is helpful in stock price prediction. The findings are very positive with excellent predictive accuracy and meteoric speed.

Keywords (separated by '-')	Reinforcement Learning - Stock Price - Stock Market Prediction - Q-learning

Stock Price Prediction Using Reinforcement Learning

Poonam Rani, Jyoti Shokeen, Anshul Singh, Anmol Singh, Sharlin Kumar, and Naman Raghuvanshi

Abstract With the availability of new data sources and advancement in marketing and financial instruments, stock market returns are a major research area. Stocks have a huge influence on today's economy. A better predictive model is extremely important in stock prediction. The aim of this paper is to investigate the positive effect of reinforcement learning on stock price prediction techniques. Q-learning has been shown to be incredibly effective in various segments, such as cloud scheduling and game automation. This paper demonstrates how the Q-learning technique is helpful in stock price prediction. The findings are very positive with excellent predictive accuracy and meteoric speed.

Keywords Reinforcement Learning · Stock Price · Stock Market Prediction · Q-learning

P. Rani (✉) · A. Singh · A. Singh · S. Kumar · N. Raghuvanshi
Department of Computer Engineering, Netaji Subhas University of Technology, Dwarka New Delhi, India
e-mail: anshuls1.co.17@nsit.net.in

A. Singh
e-mail: anmols.co.17@nsit.net.in

S. Kumar
e-mail: sharlink.co.17@nsit.net.in

N. Raghuvanshi
e-mail: namanr.co.17@nsit.net.in

J. Shokeen
Department of Computer Science and Engineering UIET, Maharshi Dayanand University, Rohtak, Haryana, India
e-mail: jyotishokeen.rs.uiet@mdurohtak.ac.in

A. Khanna et al. (eds.), *International Conference on Innovative Computing and Communications*, Advances in Intelligent Systems and Computing 1388,
https://doi.org/10.1007/978-981-16-2597-8_6

1 Introduction

AQ1 Today, stock exchanges are one of the most profitable fields of stock trading in the business world. As the industry is becoming prevalent, there is an overwhelming need for better and faster prediction models as stocks directly impact the future of the business as well as the future of investors. The stock exchange can be thought of as a game where you have to sell, buy, and hold stocks at the right time by analyzing the financial market and in some cases with your gut. But still, no one can predict the future which makes this game extremely hard and risky. How accurately you can predict the future of the company will decide your victory, and in this case, victory could mean millions of dollars earned in no time, but at the same time loss would mean millions of dollars lost in no time. This is why there is an escalating need for an efficient stock price prediction model. Consequently, stock prediction is a hot topic in the research and development sector.

Many stock prediction models, including Convolutional Neural Network (CNN), Recurrent Neural Networks (RNN), and Long short-term memory (LSTM) are based on neural networks [2]. Reinforcement learning is one of the machine learning techniques that deals with how the agent performs actions to maximize cumulative reward in an environment [11]. This technique functions on the reward and punishment policy. It implies that the model penalizes each time it does not work for the solution, and gives reward if the action turns into victory. Reinforcement learning offers simple solutions to several complex problems that other supervised and unsupervised machine learning algorithms. AQ2

In the past decade, stock traders had to rely on different software intelligence systems to reach trading decisions. Lately, with the evolution of artificial intelligence networks, this field has completely changed and has experienced a huge reform. Apart from reinforcement learning, there are numerous machine learning algorithms that can be used efficiently in stock prediction such as CNNs, RNNs, and LSTM with sliding windows. We employ reinforcement learning for stock price prediction in this paper because reinforcement learning allows the use of market signals to create profitable trading strategies in a trading context.

The subsequent sections are organized as follows: Sect. 2 introduces some recent works related to this area in the literature. Sections 3 and 4 introduce the reinforcement learning approach and stock price prediction, respectively. Section 5 defines the methodology used in the paper. Section 6 discusses the experimental work and the results. Lastly, Sect. 7 concludes the paper.

2 Related Works

Parmar et al. [4] also worked in this direction of stock market prediction to predict future values of financial stocks. They used linear regression and LSTM to propose the model. Linear regression is used for predicting continuous values by reducing the error function, i.e., gradient descent. They used LSTM for prediction on a large amount of data. However, LSTM needs a huge amount of historical data for training purposes to get good accuracy. Compared to Q-learning, LSTM requires more memory for the training dataset. Also, Q-learning provides a sense of random actions to be taken like humans, which is not possible in LSTMs. LSTMs can just predict stock prices but cannot take actions like buy, sell, or hold according to the predictions.

In paper [7], the authors used Support Vector Machine (SVM) as the classifier to know about the action to perform. They believe that SVM is one of the most suitable algorithms for time series prediction. Li et al. [3] employed deep reinforcement learning for stock transaction strategy. They claimed that their algorithm is more intelligent than traditional algorithms because of fast adaption and response for changes. However, their approach is not feasible for large datasets.

3 Reinforcement Learning

Reinforcement learning is one of the machine learning areas which is concerned with how the algorithms decide to take actions in an environment to augment the cumulative reward. Unlike supervised algorithms, reinforcement learning does not need a labeled input/output dataset. Instead, it focuses on finding the balance between exploration and exploitation. Reinforcement learning works similar to the human brain. We take action based on some past experience and our intuition. We assess the result and reward ourselves if it turns out to be profitable and learn that this is a viable action. But in case of loss, we penalize ourselves and try a different way to solve the problem. This is how reinforcement learning works: it grants rewards if the algorithm's action results in a win and penalizes if it loses. It learns each time it makes a prediction. Figure 1 portrays the functioning of reinforcement learning.

Q-learning is an off-policy reinforcement learning approach that aims to get the best action from the given current state. Q-learning is treated as off-policy because it does not need any policy and the Q-learning function learns from actions that are outside the current policy such as taking a random action. "Q" stands for quality AQ3 in Q-learning. Quality indicates how productive a given action is in gaining some future reward. The goal of Q-learning is to maximize the total reward [1]. Li et al. [3] applied deep reinforcement learning in stock forecasting.

4 Stock Price Prediction

In the financial world, forecasting stock prices is an essential goal. A fairly accurate forecast has the potential to yield high financial benefits and protect against market risks. An efficient stock price prediction system can result in a huge amount of profit in the future. The theory of an efficient market implies that stock prices represent all the information currently available, and any price adjustments that are not based on newly

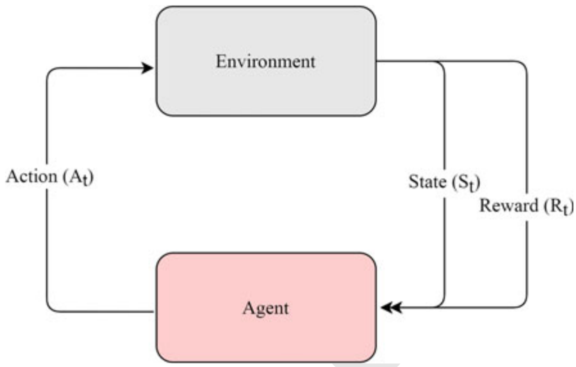

Fig. 1 Reinforcement learning

released information are therefore potentially unpredictable. Others disagree and those with this perspective have countless techniques and technologies that allegedly allow them to gain information on future prices. However, due to the uncertainty and unpredictable nature of the markets and the many undecidable, non-stationary stochastic variables involved, forecasting stock prices is not a simple task. Nowadays, social network analysis is useful in predicting stock prices [6, 8]. On predicting the stock prices, the users can use the recommender systems to buy or sell the stocks [5, 9, 10].

The historical trends of financial time series have been analyzed by several scholars from different areas, and different methods for forecasting stock prices have been proposed. Most of these methods involve careful selection of input variables to achieve promising results, developing a predictive model with skilled financial expertise, and introducing different statistical methods for arbitrage analysis, making it impossible for individuals outside the financial sector to use these methods to forecast stock prices.

5 Methodology

The method that we used in this paper for stock price prediction is Q-learning. We are here first creating an environment and an agent. The term environment in reinforcement learning is referred to as the task, i.e., stock price prediction and the agent refers to the algorithm used to solve that particular task. Hence, the driver program just initiates the needed environment and agents which are given as input to the algorithms which return predictions in values. This part of the algorithm is responsible to calculate the gradient descent or the algorithm which eventually talks about the accuracy of the algorithm.

We incorporate two additional functions, i.e., the reset function and the step function. The reset function's task is to bring back the pointer to zero, i.e., start of the time, where the cash in hand is maximum and the investment is zero. The step function takes in action as the input and performs action accordingly, i.e., it buys the stock, moves the pointer, and at the same time updates the reward, next state, and portfolio values.

Unlike the typical $Q(s, a)$, we use only the state s and ignore the action a in this stock problem.

$$Q(s, :) = W^T.s + b \tag{1}$$

where $Q(s, :)$ is the vector of Q values at state s, W is the matrix of weights, s is the state, and b is bias.

This part of the algorithm contains mainly three functions with various tasks, i.e., *_init_* function, *get_action* function, and the train function. The *_init_* function is just used to initialize the model used for training purposes. The *get_action* function takes the state as the input and accordingly decides which action to be taken, i.e., whether to buy stock, sell stocks, or remain sprawl using reinforcement learning techniques such as epsilon or greedy. Finally, the train function takes a tuple of data including current state, action, reward, next state, and done flag. It calculates the input and target values that are input to our model, where input is the state and the target is calculated as follows:

$$target = r + \gamma * max Q(s', :) \tag{2}$$

where γ is the discount factor which is used to align existing benefits and potential ones. $max Q(s', :)$ is the maximum Q value among all possible actions given state s'.

6 Experimental Work

In order to make it a real-time project, it is necessary to take recent data. API Alpha Vantage is used to get real-time data to make good predictions about a stock. Alpha vantage is a free API which is used to provide real-time data of the stock. We take the stock prices of three companies: Apple, Microsoft, and International Business Machines (IBM) for the duration of January 2012–November 2020. The dataset is divided into a 60:40 ratio as training set and testing set, respectively.

The first step to every reinforcement technique is to decide on an action in the beginning. It is assumed that before making any decision, the algorithm must know the answers to some questions such as

- Do I even have enough cash to buy?
- Considering the current state of my portfolio and the existing price of the shares in the market, is it worth selling them?

After answering all these questions, the next step is to decide the action to be taken. There are three actions: buy, sell, or hold. Reinforcement learning algorithms evaluate the action and make the next decision accordingly. Reward is the difference

Table 1 Initial model configurations

Parameter	Value
Episodes	200
Initial investment	20000
ε	1.0

Table 2 Parameters chosen for experiments

Parameter	Value
Exploration rate	1.0
Epsilon decay	0.996
Discount factor	0.94

Table 3 Performance results

Performance parameter	Reward
Average reward	38356.61
Minimum reward	23091.67
Maximum reward	54582.06

between portfolio values of recent time steps and previous time steps. The algorithm computes the portfolio value as follows:

$$value = S^T.P + C \tag{3}$$

where S is the vector of shares owned, P is the vector of share prices, and C is the cash.

Epsilon decay is the value of ε to learn and act optimally in the life of an agent. We experimented on different values of $epsilon\_decay$ and γ to find their best value. Table 1 depicts the initial configurations set in the proposed model.

Table 2 defines the parameters tuned for the experiments and Table 3 shows the performance of the model based on the selected parameters. Figure 2 depicts the results of rewards in respect of $epsilon\_decay$ for $\gamma = 0.95$, $\varepsilon = 1.0$, and $\varepsilon\_min = 0.01$. The average reward is 32697.13 and the best value for $epsilon\_decay$ in terms of reward is 0.996. Figure 3 depicts the results of rewards in respect of γ with exploration rate $\varepsilon = 1.0$, $\varepsilon\_min = 0.01$, and $\varepsilon\_decay = 0.996$. The average reward is 81588.52 and the best value for γ in terms of reward is 0.94.

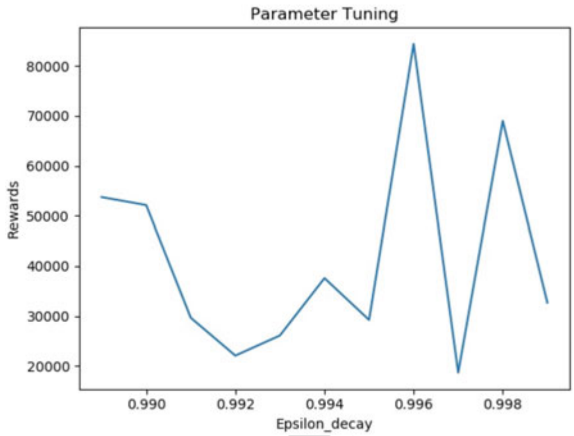

Fig. 2 Test results of rewards versus *epsilon_decay*

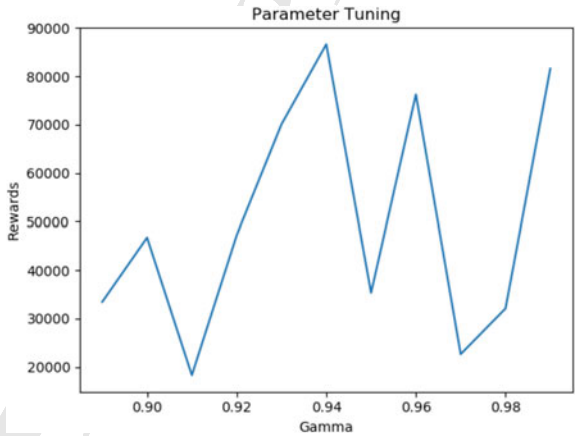

Fig. 3 Test results of rewards versus γ

7 Conclusion

In this paper, we used reinforcement learning for stock price prediction. Also, we found its performance which allows us to use this model as a base of further study in this domain. Reinforcement learning is better than other learning algorithms as it learns from current situations and also makes faster adaptive changes.

References

1. C. Jin, Z. Allen-Zhu, S. Bubeck, M.I. Jordan, Is Q-learning provably efficient? in *Advances in neural information processing systems* (2018), pp. 4863–4873
2. C.K.S. Leung, R.K. MacKinnon, Y. Wang, A machine learning approach for stock price prediction, in *Proceedings of the 18th International Database Engineering & Applications Symposium* (2014), pp. 274–277
3. Y. Li, P. Ni, V. Chang, Application of deep reinforcement learning in stock trading strategies and stock forecasting. Computing 1–18 (2019)
4. I. Parmar, N. Agarwal, S. Saxena, R. Arora, S. Gupta, H. Dhiman, L. Chouhan, Stock market prediction using machine learning, in *2018 First International Conference on Secure Cyber Computing and Communication (ICSCCC)* (IEEE, 2018), pp. 574–576
5. P. Rani, J. Shokeen, D. Mullick, Recommendations using modified k-means clustering and voting theory. Int. J. Comput. Sci. Mobile Comput. **6**(6), 143–148 (2017)
6. P. Rani, D.K. Tayal, M. Bhatia, SNA using user experience, in *2019 International Conference on Machine Learning, Big Data, Cloud and Parallel Computing (COMITCon)* (IEEE, 2019), pp. 125–128
7. V.K.S. Reddy, Stock market prediction using machine learning. Int. Res. J. Eng. Technol. **5**(10)(2018)
8. J. Shokeen, C. Rana, Social recommender systems: techniques, domains, metrics, datasets and future scope. J. Intell. Inform. Syst. **54**, 633–667 (2019). https://doi.org/10.1007/s10844-019-00578-5
9. J. Shokeen, C. Rana, A study on features of social recommender systems. Artif. Intell. Rev. **53**(2), 965–988 (2020)
10. J. Shokeen, C. Rana, P. Rani, A trust-based approach to extract social relationships for recommendation, in *Data Analytics and Management* (Springer, 2020), pp. 51–58
11. R.S. Sutton, A.G. Barto, *Reinforcement Learning: An Introduction* (MIT press, 2018)

Author Queries

Chapter 6

Query Refs.	Details Required	Author's response
AQ1	Please check and confirm if the author names and initials are correct.	
AQ2	Please check the sentence "Reinforcement learning offers ..." for clarity.	
AQ3	Please approve the edits made to the sentence "Q-learning is treated ...".	

MARKED PROOF

Please correct and return this set

Please use the proof correction marks shown below for all alterations and corrections. If you wish to return your proof by fax you should ensure that all amendments are written clearly in dark ink and are made well within the page margins.

Instruction to printer	*Textual mark*	*Marginal mark*
Leave unchanged	• • • under matter to remain	
Insert in text the matter indicated in the margin		New matter followed by or
Delete	through single character, rule or underline or through all characters to be deleted	or
Substitute character or substitute part of one or more word(s)	through letter or through characters	new character or new characters
Change to italics	under matter to be changed	
Change to capitals	under matter to be changed	
Change to small capitals	under matter to be changed	
Change to bold type	under matter to be changed	
Change to bold italic	under matter to be changed	
Change to lower case	Encircle matter to be changed	
Change italic to upright type	(As above)	
Change bold to non-bold type	(As above)	
Insert ‘superior’ character	through character or where required	or under character e.g. or
Insert ‘inferior’ character	(As above)	over character e.g.
Insert full stop	(As above)	⊙
Insert comma	(As above)	,
Insert single quotation marks	(As above)	or and/or or
Insert double quotation marks	(As above)	or and/or or
Insert hyphen	(As above)	
Start new paragraph		
No new paragraph		
Transpose		
Close up	linking characters	
Insert or substitute space between characters or words	through character or where required	
Reduce space between characters or words	between characters or words affected	