Transdisciplinary Engineering for Complex Socio-technical Systems
K. Hiekata et al. (Eds.)

doi:10.3233/ATDE190169

Optimization Method for Labor Control in Net Shopping Logistics

Kazuhiro KOIKE[1] and Yingsha YANG[1]
ASKUL Corporation, Japan

Abstract. In the retail industry, to increase sales and the number of users, it is common practice to regularly or irregularly carry out sales promotion measures such as discount sales and point return. The fluctuation in demand due to such sales promotion measures is amplified compared to when it is not implemented. This wave propagates along with the lead time (difference between the order receipt date and the designated delivery date), since the order is placed on the EC site and shipped on the designated delivery date, unlike in the real store sales. This wave of late-propagating fluctuations not only has a major impact on the scheduling of workers shipping at the warehouse, but also has a negative impact on the business. In other words, excessive staffing increases shipping costs, and staffing shortages cause delivery delays, resulting in lower customer satisfaction. Therefore, we propose a method to trade-off between customer satisfaction and shipping cost by formulating an appropriate shipping plan that anticipates fluctuations in demand due to sales promotion measures.

Keywords. Analysis and Engineering, Supply Chain and Logistics, Cost Modeling, Decision Supporting Tools and Methods

Introduction

Generally, Bullwhip effect (notated as BE below) is known as one of the problems that may occur in the supply chain. This is a phenomenon in which demand propagates from downstream to upstream while expanding as a result of demand forecasting and decision making downstream of the supply chain, and it was first recognized in 1958 [1]. Since this phenomenon leads to excess inventory and shortages, the generation mechanism and control methods have been studied for many years by Lee et al [2]. Factors for BE include price list, order frequency, return policy, frequency and depth of price sales measures, degree of information sharing, demand forecasting method, allocation rule at the time of stockout etc., are listed [3].

The appearance of BE is often reproduced by Beer Game (notated as BG below). BG is a simulation game invented at MIT in the 1960s. In the beer supply chain connected in series, four players, namely retailers, wholesalers, distributors and manufacturers, compete for cost minimization within a certain period of time.

It had already been pointed out by Lee et al that excessive promotional measures are a factor in BE. In Internet shopping, it is structurally integrated without intentionally taking excessive sales promotion measures. This is because moving Atom with material and weight is not as easy as moving Bit.

[1] Corresponding Author, Email: kazuhiro.koike@askul.com

The problems that can arise due to differences between cyber and physical can not be reproduced with the original BG. Therefore, in this research, we newly set up a simulation environment in which the rules of BG have been changed for Internet shopping. Under such circumstances, in order to trade off sales promotion measures and shipping costs in the online shopping distribution process, we examined and evaluated a method using deep reinforcement learning.

1. Literature Review

The process of BG is known as MDP (Markov Decision Process), and the information that can be observed is between adjacent players. It is POMDP (Partially Observable Markov Decision Process) because it is only the order, the exchange of goods, and its own inventory level. Each player try to minimize costs from observable partial information. The observation space and the action space are large and complex problems because they deal with nonstationary time series. The DQN (deep Q-network) [4] proposed by Mnih et al is promising as a way to overcome such complex problems. The deep reinforcement learning approach of BG includes, for example, [5] and [6], and its effects have been reported.

2. Model

2.1. Overview

In the internet shopping logistic process, there are differences in characteristics between the process that is completed in cyber space such as EC site and net transaction, and the process that is performed in physical space such as distribution warehouse. For example, in cyber space, 100 products are just numerical data (Bit), and there is no physical restriction on making 100 products 10 times by bullish sales promotion measures. On the other hand, in physical space, 100 items are real materials with volume and weight (Atom), and are greatly affected by physical constraints such as the capacity of the warehouse and the maximum number of shipments.

In BG, four players connected in series, namely Retailer, Wholesaler, Distributor, and Manufacturer, participate in the game. However, in this research, in order to focus on the above-mentioned problems specific to online shopping, the player is only Retailer, and only customer orders and shipments are handled. Orders are limited to the number reserved for inventory and do not consider out of stock. The demand from the customer shall be increased by the sales promotion measures. The purpose of the Player is to ship the number of ordered items within the maximum shipping number and to minimize the labor cost of the shipping worker. Products ordered from the EC site are basically shipped the next day and delivered to customers, but there is a limit to the number of products that can be shipped in one day, and it can not be shipped beyond that. The amount exceeding the maximum number of shipments results in lowering customer satisfaction by being shipped after the next day. On the contrary, if there is only a demand lower than the shipping plan, it will be a surplus cost of shipping workers.

Assuming that the number of demand is (d) and the number of planned shipments is (s) and the difference number is (z = d - s). In the case of (d) is greater than (s), (z) will be shipped after the next day. It is multiplied by the cost coefficient (Cs) to express

customer satisfaction as a cost. If (d) is less than (s), the surplus cost will be the multiplying (z) by the cost coefficient (Cl). It is assumed that the degree of promotion (x), the holiday flag (h), the day of the week (y), and the regular sales promotion day flag (f) are given as observable parameters. The problem is finding the appropriate number of shipping plans (s) that does not exceed the maximum shipping volume with minimul costs.

As an index for evaluating the solution obtained, the negative value of cost is considered as reward (- cost), and our goal is to maximize the reward. Also, let BEI (Bullwhip Effect Index) be the variance var (s) of the number of shipment plans (s) of one episode (1,000 steps) divided by the variance var (d) of demand (d). Control BEI to equalize the shipping plan to the demand. Also, let FR (Fill Rate) be the average of one episode of the ratio of shipments to demand. In other words, it indicates how much can be shipped to demand. State variables are the observable parameters plus the number of planned shipments (s), demand (d), and the difference between the demand and the number of planned shipments (z). This state variable evaluation is modeled by DQN, and reinforcement learning by interaction between agent and environment is performed (Figure 1). It is implemented with OpenAI Gym, a reinforcement learning development environment.

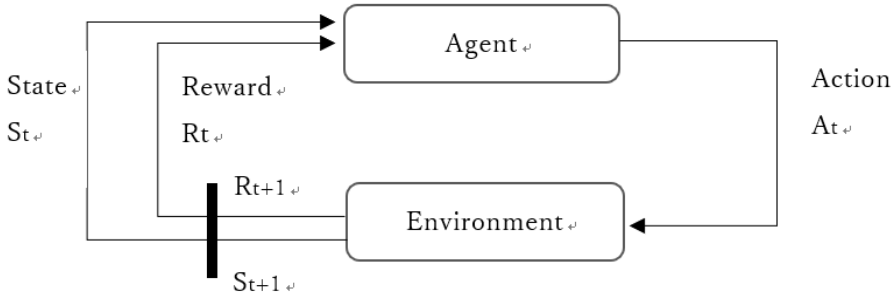

Figure 1. Interaction between agent and environment in reinforcement learning.

2.2. *Problem Definition*

Suppose that there is fluctuation in the number of demands as shown in Figure 2. The vertical axis is the number of demand, and the horizontal axis is the day. The three peaks are the days when the sales promotion measures were implemented, and increase about three times the usual. Items ordered will be shipped the next day or later. Consider the four-pattern shipping plan in this case.

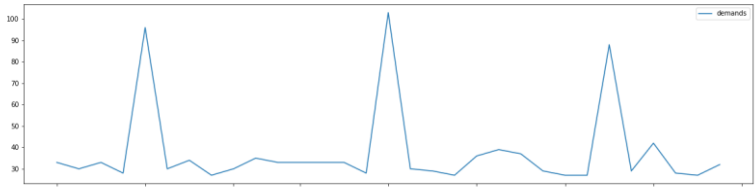

Figure 2. fluctuation in the number of demands by the sales promotion measures.

1. Maximize customer satisfaction

The shortest delivery plan from the order is as shown in Figure 3. This is the plan with the highest customer satisfaction, but the number of shipments per day is limited. If the maximum number is 55, you can not achieve more than that.

2. Simplest labor management
   Figure 4 is a shipping plan to keep the number of shipments constant. Labor management is the simplest, but the lead time from order acceptance to shipping becomes longer and customer satisfaction declines.
3. Looks good but difficult labor management
   The shipment plan will be shipped the maximum number of shipments the day after the sales promotion measures peak, and will be gradually reduced thereafter (Figure 5). Although this plan looks the best, labor management is difficult.
4. Better shipment plan among four
   From the next day of the sales promotion measures peak, we will ship within the maximum number of shipments for several days, and thereafter, we will make a normal shipment plan (Figure 6). Of the four patterns, it is the most balanced shipping plan. In this study, we aim to make this form by reinforcement learning.

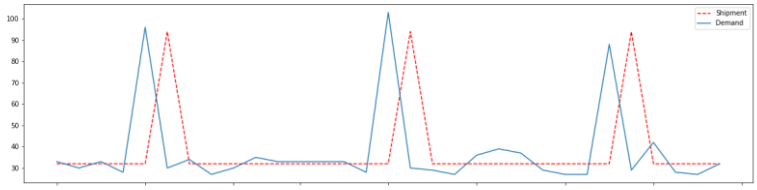

Figure 3. Maximize customer satisfaction.

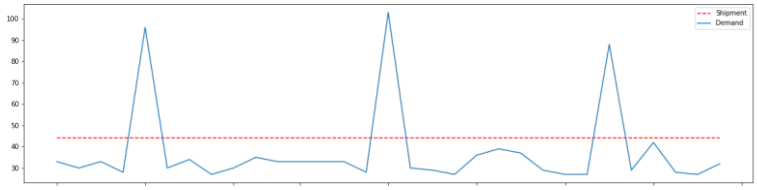

Figure 4. Simplest laabor management.

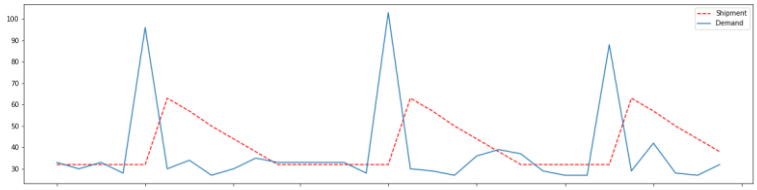

Figure 5. Looks good but difficult labor management.

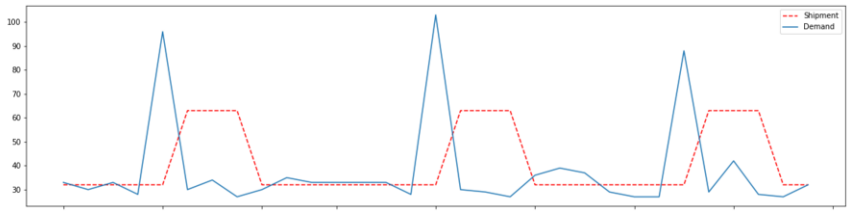

Figure 6. Better shipment plan among four.

2.3. Environment Definition

2.3.1. State values

Equation 1 defines the observable state variable o_t at time step t. Equation 2 defines ho_t which stores state variables of all time steps of one episode as historical observation.

$$o_t = [h, y, f, x, d, z, s] \tag{1}$$

$$ho_t = [o_1, o_2, o_3, \ldots . o_t] \tag{2}$$

h : Holiday flag. Saturday and Sunday have 10 other days as 1.
y : Day of the week. 1:Tue, 2:Wed, 3:Thr, 4: Fri, 5:Sat, 6:Sun, 7:Mon
f : Regular promotion day flag. Let sales promotion day be 10 otherwise 1.
x : Degree of sales promotion measures.
d : Number of demand.
z : Difference between d and s.
s : Number of shipping plans (numbers predicted by agents learned by DQN)

2.3.2. Action space

Agent's action is the number of shipments of items to the customer. The Action space is a set of discrete values from 1 to max_shipment per day.

2.3.3. Reward

As shown in equation (3)(4)(5), the cost is calculated according to the difference (z) between the number of demand (d) and the number of planned shipments (s), and the value with negative cost is used as the reward.

$$\text{cost} = z \ \ C_s \tag{3}$$

$$\text{cost} = \text{abs}(z) \ \ C_l \tag{4}$$

$$\text{reward} = -(\text{cost}) \tag{5}$$

C_s : Customer Satisfaction Cost Factor
C_l : Shipper cost factor
Max_shipment : Maximum shipments per day

2.3.4. Goal condition

No goal conditions are set. One episode has 1,000 steps, and it is evaluated by the sum of rewards acquired in one episode.

2.3.5. Demand Forecast

Demand is generated based on the distribution of demand for a certain product group for a period of less than 20 months.

2.4. Algorithm

The ε-greedy algorithm was adopted as a method to balance the accumulation and utilization of the experience obtained from the result of the Action, and the DNN was adopted for the state evaluation (Figure 7). The upper limit of learning was 50,000 steps, and one episode was learned in 1,000 steps. I learned 50 episodes. Figure 8 shows how the reward increases as the agent's learning progresses.

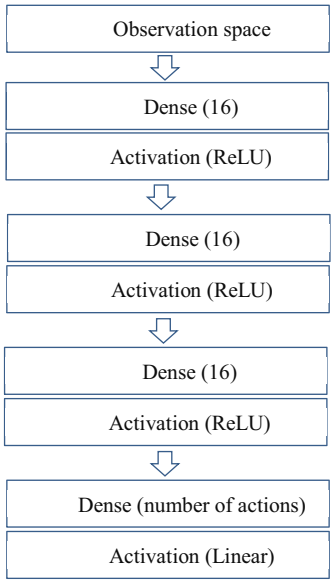

Figure 7. DNN Model.

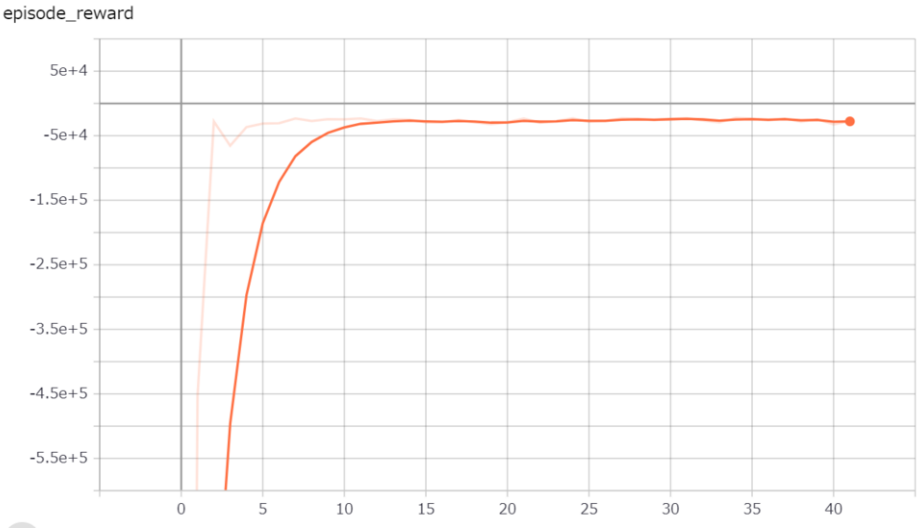

Figure 8. Episode-Reward progress.

3. Case Study & Discussions

3.1. Data Overview

We tested 50 episodes using the learned agent model. The time step for one episode is 1,000. Table 1 shows the acquired reward for each episode, BEI, FR, and the top 5 acquired rewards for one episode. Figure 9 is last 90 time steps of demands and shipments graph of top 5, and Figure 10 is last 30 time steps graph.

By adjusting the values of Cs and Cl, the Agent was able to learn to output a form close to the originally intended shipping plan waveform (Figure 6). However, as shown in Figure 10, Episode # 42, there is a pile of excess shipping plans between the high peaks and the peaks, and there is room for improvement.

Table 1. The acquired reward, BEI and FR for each episode. The top 5 reward episodes.

Episode	Reward	BEI	FR
Episode 49	-20739.9	0.372618	1.362783
Episode 22	-20725.9	0.372388	1.362919
Episode 42	-20691.5	0.376941	1.368707
Episode 45	-20690.2	0.374528	1.361611
Episode 18	-20689.2	0.375318	1.361265

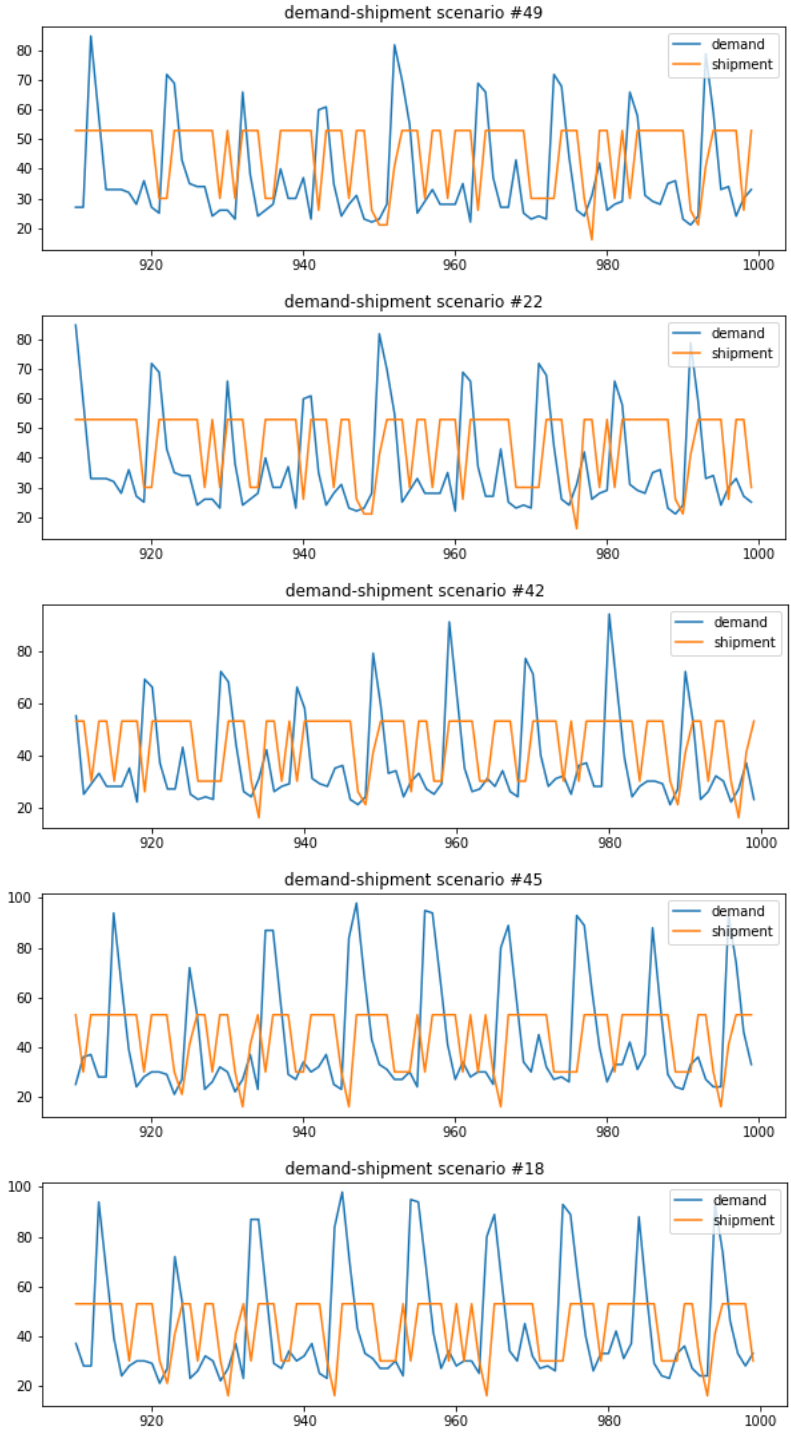

Figure 9. The top 5 episodes of demand-shipment with last 90 time steps.

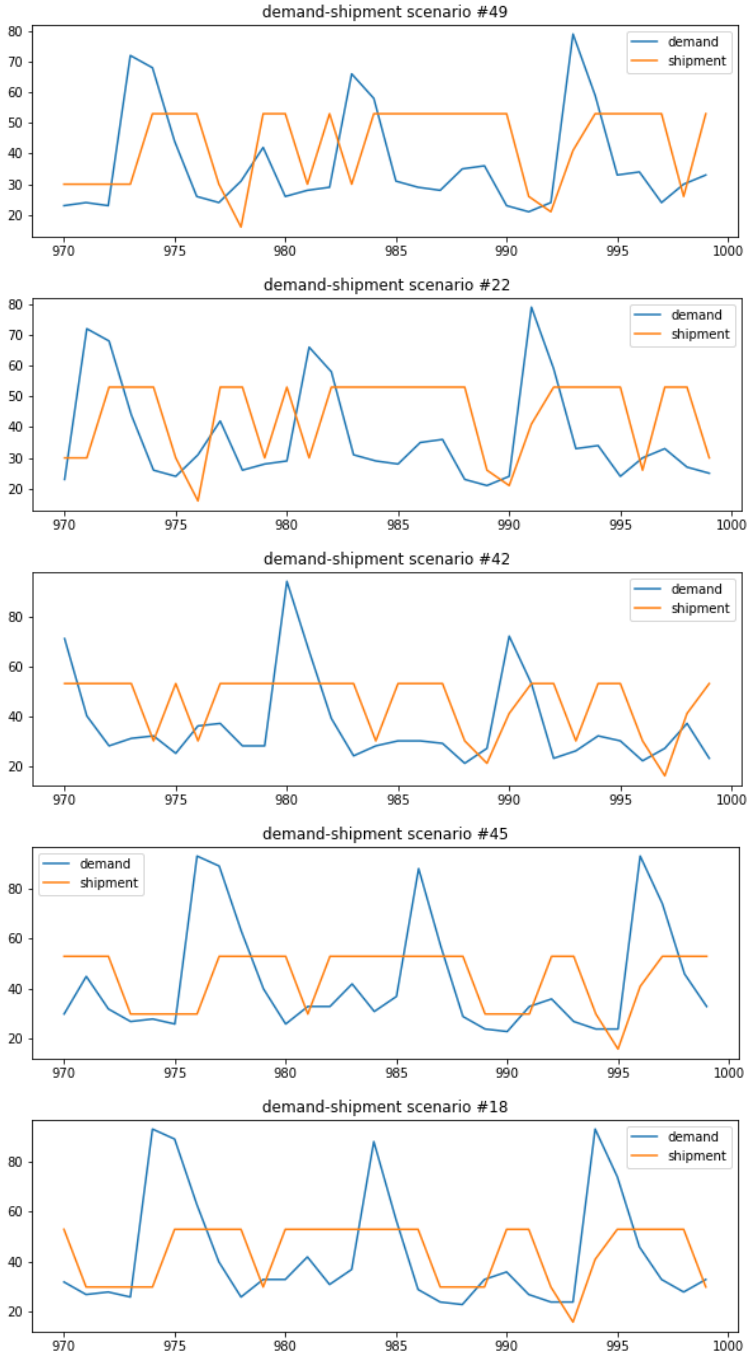

Figure 10. The top 5 episodes of demand-shipment with last 30 time steps.

4. Conclusion

The difference between the 1960s, when BG was conceived, and the current prosperity of Internet shopping, is that the flow of information in the supply chain has significantly evolved and become more efficient. Compared with that, the speed of evolution is relatively slow for physical distribution, so while cyber space such as EC site and net transaction advances digitization and data utilization by AI advances, efficiency between physical space such as warehouse and delivery is relatively delayed.

In general, in a functionally divided organization, cyber players and physical players individually define KPIs and pursue their maximization. It looks reasonable from a microscopic point of view focusing on each function, but in situations where there is a trade-off between functions, overall optimization can not be achieved. Overall optimization requires a high dimensional or meta perspective.

There are many problems that there is no teacher data in the distribution problem, and there is information that is difficult to share among players and information that is difficult to observe in the first place. Because changing facilities is expensive, trying different methods is not easy. Therefore, we considered that deep reinforcement learning that can be learned from the interaction in the virtual environment is effective. This study is the first step to show its potential.

Acknowledgement

The author would like to thank those who provided valuable advice.

References

[1] J.W. Forrester. Industrial dynamics: A major breakthrough for decision makers. *Harvard Bus. Review*, (July/August 1958) 36 3766.Y.

[2] H. L. Lee et al., Comments on "Information Distortion in a Supply Chain: The Bullwhip Effect", *Management Science*, 2004, Vol. 50 (12 supplement), pp. 1887-1893.

[3] H. L. Lee et al., Information distortion in a supply chain: The bullwhip effect, *Management Science*, 1997, Vol. 43, No. 4, pp. 546-558.

[4] V. Mnih et al., Human-level control through deep reinforcement learning, *Nature*, 2015, Vol. 518, pp. 529–533, doi 10.1038/nature14236.

[5] Taiki Fuji et al., Deep Multi-Agent Reinforcement Learning using DNN-Weight Evolution to Optimize Supply Chain Performance, *Proceedings of the 51st Hawaii International Conference on System Sciences*, 2018, Doi 10.24251/HICSS.2018.157.

[6] Afshin Oroojlooyjadid et al., A Deep Q-Network for the Beer Game: A Reinforcement Learning Algorithm to Solve Inventory Optimization Problems, arXiv:1708.05924v2, 2018.