Expert Systems With Applications 91 (2018) 150–158

Contents lists available at ScienceDirect

Expert Systems With Applications

journal homepage: www.elsevier.com/locate/eswa

Reinforcement learning approaches for specifying ordering policies of perishable inventory systems

Ahmet Kara, Ibrahim Dogan*

Erciyes University, Industrial Engineering Department, Kayseri, Turkey

ARTICLE INFO

Article history:
Received 23 May 2017
Revised 23 August 2017
Accepted 24 August 2017
Available online 30 August 2017

Keywords:
Reinforcement learning
Inventory management system
Simulation-based optimization
Ordering management
Perishable item

ABSTRACT

In this study, we deal with the inventory management system of perishable products under the random demand and deterministic lead time in order to minimize the total cost of a retailer. We investigate two different ordering policies to emphasize the importance of the age information in the perishable inventory systems using Reinforcement Learning (RL). Stock-based policy replenishes stocks according to the stock quantities, and Age-based policy considers both inventory level and the age of the items in stock. The problem considered in this article has been modeled using Reinforcement Learning and the policies are optimized using Q-learning and Sarsa algorithms. The performance of the proposed policies compared with similar policies from the literature. The experiments demonstrate that the ordering policy which takes into account the age information appears to be an acceptable policy and learning with RL provides better results when demand has high variance and products has short lifetimes.

1. Introduction

One of the concerns of the companies producing perishable products is to have an effective inventory management system to improve the customer service and provide competitive advantages in the current global marketplace. However there are a number of challenges, such as stochastic customer demand, perishable nature of the products and need to trace the age of the products along the supply chain (Kouki & Jouini, 2015). In this study, we deal with a retailer's ordering replenishment problem under stochastic customer demand, limited shelf lifetime and fixed lead time situation to find the best balance between the outdating quantity and the shortage quantity. Fresh agri-products, dairy products, pharmaceuticals and human blood are typical examples of perishable products in which the age of the products are significantly important. For example platelet, one of the blood components can be used effectively for only 5 days. If the platelet stocks cannot be managed effectively by the age and quantity information, it is possible to observe a lot of stock-outs or wastage of platelets.

Classical inventory systems which do not consider the age of the products is not suitable for perishables and perishable inventory systems require different approaches (Tekin, Gürler, & Berk, 2001). Perishable inventory policies with stochastic demand have been commonly modeled using only quantity of stocks information. With the development of technology, age-based policies using the information about the age of stocks are evaluated in detail for the perishable items (Broekmeulen & van Donselaar, 2009). Heuristics approaches are proposed in the literature due to the modeling and solution complexity of the exact models. We suggest and evaluate approximate aged-based replenishment policies which take into account both the quantity and the age of the stocks. Moreover, this paper demonstrates that reinforcement learning methods can be used to solve the inventory control problem of perishable items with uncertain customer demand.

In the area of machine learning, Reinforcement learning offers the advantage of solving the complex sequential decision-making problems based on learning from the previous knowledge. Reinforcement learning provides a dynamic learning against the changing environment and no need to pre-determined model of environment (Rana & Oliveira, 2015). A variety of problems such as job-shop scheduling (Zhang & Dietterich, 1995), revenue management optimization (Gosavi, Ozkaya, & Kahraman, 2007), inventory management in the supply chain (Jiang & Sheng, 2009) and goal-regulation in manufacturing systems (Shin, Ryu, & Jung, 2012) are formulated using reinforcement learning under non-stationary and unknown conditions. In this article, we apply reinforcement learning approaches for indicating the importance of the age information of the perishable inventory system. The essential contribution of our research is the use of the Q-learning and Sarsa algorithm based on reinforcement learning to specify the near-optimal ordering replenishment policy of perishable products with stochastic customer demand and lead time.

* Corresponding author.
E-mail addresses: ahmet.kara@erciyes.edu.tr (A. Kara), idogan@erciyes.edu.tr (I. Dogan).

http://dx.doi.org/10.1016/j.eswa.2017.08.046
0957-4174/

The paper is organized as follows. In the Section 2, we provide related literature on the replenishment learning approach and the inventory management system of perishable items. Section 3 generally introduces the replenishment learning model. In the Section 4, the inventory replenishment models and proposed algorithms based on reinforcement learning are described. Details of the conducted experiments are explained in Section 5. The results of computational experiments, discussion and conclusion are presented in Section 6.

2. Literature review

In recent years, comprehensive studies related to the integration of reinforcement learning model and inventory control in the supply chain have been presented in the literature. Reinforcement learning methods have been implemented to specify near-optimal ordering policies in the entire supply chain.

Kim, Jun, Baek, Smith, and Kim (2005) suggested two inventory-control models with a nonstationary and unknown customer demand consists of centralized model and decentralized model. They employed a reinforcement learning approach called the action-value method for satisfying a predetermined target service level during the lead time. Chaharsooghi, Heydari, and Zegordi (2008) presented an inventory control system with the case of uncertain lead-times and uncertain customer demand to determine ordering policies of each echelon in the supply chain. The proposed model is formulated as a reinforcement learning approach so as to minimizing the total inventory cost including holding cost and backorder cost.

Dogan and Güner (2015) concentrated on the ordering and pricing problems of the supply chain with a multi-retailer environment and formulated the problem using a reinforcement learning method called Q-learning. Kim, Kwon, and Baek (2008) considered a learning model which updating the reward values of each action at each decision period for the supply chain with two-echelon under unstable demand environment. A similar learning model called asynchronous action-reward learning has been applied by Kwak, Choi, Kim, and Kwon (2009). It is discussed by Rana and Oliveira (2015) that how reinforcement learning may be implemented to estimate accurately the optimal pricing policies of perishable items and services utilizing.

Recently in the literature, reinforced learning has been employed to solve different problems. Almahdi and Yang (2017) suggested the recurrent reinforcement learning method in order to optimize financial investments. Zhou, Hao, and Duval (2016) use the reinforcement learning based local search (RLS) approach to solve grouping problems that are NP-hard. A Q-learning algorithm based on reinforcement learning for the train rescheduling problem has been addressed by Šemrov, Marsetič, Žura, Todorovski, and Srdic (2016). Okdinawati, Simatupang, and Sunitiyoso (2017) applied reinforcement learning to formulate to reduce transportation costs, raise visibility and develop agility in Collaborative Transportation Management. Mannion, Mason, Devlin, Duggan, and Howley (2016) proposed a Multi-Agent Reinforcement Learning (MARL) approach to a dispatch problem.

In this study we handle the issue of inventory management of perishable product taking into account stochastic customer demand and fixed lead time. The detailed reviews have been published for perishable inventory management such as Bakker, Riezebos, and Teunter (2012), Beliën and Forcé (2012), Karaesmen, Scheller-Wolf, and Deniz (2011). In order to obtain the near-optimal inventory ordering policy of perishable items are studied with many different approaches. Haijema, van der Wal, and van Dijk (2007) provided a dynamic programming and simulation-based approach to manage the platelet inventory control at the blood bank. A Markov Decision Process is formulated to model the ordering policies of pharmaceutical items by Ana, Ivy, and King (2008). Hemmelmayr, Doerner, Hartl, and Savelsbergh (2010) proposed two solution model consist of an integer programming and a Variable Neighborhood Search approach. Our study is formulated using reinforcement learning to determine near-optimal inventory control policies of perishable products.

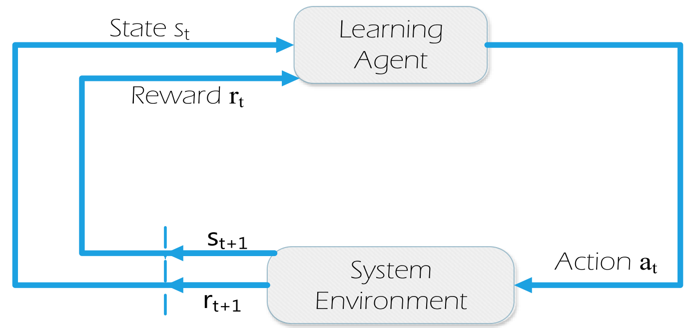

Fig. 1. The interaction of agent and environment (Sutton & Barto, 1998).

With respect to the supply chain inventory management, different models have been investigated using reinforcement learning. Pontrandolfo, Gosavi, Okogbaa, and Das (2002) considered production and distribution functions of global supply chain with multiple stages assuming a single item. The near-optimal procurement and distribution policy has been derived by a reinforcement learning model called Semi-Markov Average Reward technique (SMART). Kwon, Kim, Jun, and Lee (2008) focused on the solution of two significant situations consist of a non-stationary customer demand and a large state space. They presented a case-based myopic reinforcement learning algorithm for resolving inventory management problem of the supply chain with two-echelons. Jiang and Sheng (2009) studied on inventory management of a multi-agent supply chain structure. It has been suggested in Sun and Zhao (2012) that Q-learning is modeled as a reinforcement learning approach for specifying ordering policies of the supply chain with five stages.

3. Reinforcement learning

Reinforcement learning is a stochastic dynamic programing approach that concentrates on the concept of trial-and-error learning through the interaction of a decision-maker called agent and a dynamic environment. The structure of this approach generally consists of four basic elements: agent, state, action and reward (Das, Gosavi, Mahadevan, & Marchalleck, 1999). Fig. 1 depicts that agent-environment interaction in reinforcement learning approaches.

In each period, the learning concept in a reinforcement learning models takes place as below. When the system has a state s_t in time period t, agent chooses an action a_t among possible actions. The environment brings about a numerical reward r_{t+1} for a chosen action. The learning agent in each time period chooses an action with either the highest reward value based on its past experiences (exploitation) or a random action (exploration). The aim is to specify the optimal state-action pair that optimizes the long-run reward.

The basic issue of reinforcement learning is to resolve the problem of balancing the exploitation and exploration (Sutton & Barto, 1998). Initially, the learning agent frequently prefers the exploration strategy because it does not have enough details regarding environment. In the following phase, the agent mostly exploits the current information and the action with the highest reward value is preferred. Methods of action selection such as E3 technique (Kearns & Singh, 2002), the external-source technique (Smart & Kaelbling, 2000) and search-then-converge tech-