Situation reactive approach to Vendor Managed Inventory problem

Choonjong Kwak [a], Jin Sung Choi [b], Chang Ouk Kim [b,*], Ick-Hyun Kwon [c]

[a] Research Evaluation and Planning Division, Korea Aerospace Research Institute, 45 Eoeun-Dong, Yuseong-Gu, Daejeon 305-333, Republic of Korea
[b] Department of Information and Industrial Engineering, Yonsei University, Seoul 120-749, Republic of Korea
[c] Department of Systems Management Engineering, Inje University, Gimhae, Gyeongnam 621-749, Republic of Korea

ARTICLE INFO

ABSTRACT

In this research, we deal with VMI (Vendor Managed Inventory) problem where one supplier is responsible for managing a retailer's inventory under unstable customer demand situation. To cope with the nonstationary demand situation, we develop a retrospective action-reward learning model, a kind of reinforcement learning techniques, which is faster in learning than conventional action-reward learning and more suitable to apply to the control domain where rewards for actions vary over time. The learning model enables the inventory control to become situation reactive in the sense that replenishment quantity for the retailer is automatically adjusted at each period by adapting to the change in customer demand. The replenishment quantity is a function of compensation factor that has an effect of increasing or decreasing the replenishment amount. At each replenishment period, a cost-minimizing compensation factor value is chosen in the candidate set. A simulation based experiment gave us encouraging results for the new approach.

© 2008 Elsevier Ltd. All rights reserved.

1. Introduction

Fierce competition in today's global markets, the introduction of products with shorter life cycles, and the heightened expectations of customers have forced business enterprises to focus attention on and invest in their supply chain management (Simchi-Levi, Kaminsky, & Simchi-Levi, 2000). It is an effective means of the enterprises to improve their service levels for customers at minimum costs. One of key factors to improve their service levels is to efficiently manage the inventory level of each participant within supply chains.

Traditional "order-and-supply" based inventory control policy suffers from inordinate amount of surplus stocks at suppliers. In general, the suppliers do not know the order quantities of retailers in advance and have to maintain more safety stocks than they actually need for on-time replenishment. This causes the magnification of demand fluctuations as orders move up to upstream sites in a supply chain, which is called "bullwhip effect" (Lee, Padmanabhan, & Whang, 1997). To resolve the problem, VMI (Vendor Managed Inventory) model has been developed in industry.

VMI is a successful inventory control model for a two-stage supply chain in which a supplier directly manages the inventory level of a retailer (Achabal, Mcintyre, Smith, & Kalyanam, 2000). Within the VMI model, the retailer provides the supplier with information

on its sales and inventory level and the supplier determines the replenishment quantity at each period based on the information. Throughout the VMI model, the supplier can set up efficient replenishment plans, while the retailer can receive appropriate amounts of replenishment on time (Kaipia, Holmstrom, & Tanskanen, 2002; Lee, So, & Tang, 2000).

Customer demands have recently become more and more unstable with the widespread introduction of e-commerce, because they easily fluctuate even with minor price changes on-line. The advent of products with a variety of qualities and functions is another reason of instability in customer demands, which, in turn, increases the uncertainty of demand forecasting followed by higher inventory costs due to unnecessary inventory surplus or shortage.

In fact, inventory control has been studied for several decades for cost savings of enterprises (Axsäter, 2000, 2001; Moinzadeh, 2002; Zipkin, 2000). They have tried to maintain appropriate inventory levels to cope with stochastic customer demands and to boost their image through customer satisfaction. However, most theoretical inventory models require that the statistical characteristics of customer demand is known or can be estimated through sophisticated time series models when customer demands show nonstationary behaviors. These prerequisites are not practical from the aspects of the analysis time and efforts, especially when the supplier deals with hundreds of different items and the most of their demands fluctuate over time differently. As a consequence, the importance of situation reactive models has surfaced with the necessity of adaptively controlling the parameters of inventory

control models according to the change in customer demand (Alström & Madsen, 1996; Gavirneni & Tayur, 2001; Graves, 1999).

Action-reward learning is considered as one of the reinforcement learning techniques. The action-reward learning progressively finds the best among several possible actions in a non-static environment (Sutton & Barto, 1998) through exploitation and exploration. The basic principle of action-reward learning is as follows. When an agent chooses a certain action, the state of the non-static environment changes and the reward for the action is also determined. The reward is a numerical value that is the input to the performance measure of the action. Through the repetitive process of applying actions, the agent continuously updates the performance measures of all actions and can choose the best action based on the updated performance measures. This conventional action-reward learning generally has a trade-off between exploitation and exploration. An exploitation is to choose an action with the best value of performance measure and apply it to the non-static control system, while an exploration is to choose an action with immature learning to boost the reliability of the performance measure of the chosen action.

Kim, Jun, Baek, Smith, and Kim (2005) proposed an adaptive inventory control model for supply chains with unstable customer demands by applying the conventional action-reward learning method. The model dealt with an inventory control problem where decision variable is order interval between a supplier and a retailer. Replenishment quantity is assumed to be fixed. They proposed a method that controls both supplier's safety lead time and retailer's safety stock adaptively according to the variation of customer demand stream. The objective of the model is to satisfy the target service level predefined for each retailer. In their approach, action is probabilistically selected in order to balance the exploitation and exploration. However, the probabilistic action selection rule brings about a problem that its learning rate is getting slow with many explorations as the number of actions increases. Thus, it takes a very long time to find a good decision policy due to exploration, particularly in on-line learning.

In this paper, we propose a situation reactive VMI approach that adapts replenishment quantity over time according to the changes in customer demand stream. To cope with the nonstationary demand situation, we develop retrospective action-reward learning model that is faster in learning than the conventional action-reward learning and more suitable to apply to the control domain where rewards for actions vary over time. The retrospective analysis based model improves the learning rate of action-reward learning by eliminating the exploration.

The objective function of the inventory control is to minimize the long run average of inventory shortage and holding costs that are incurred at each replenishment period. This approach does not assume that customer demand process does follow a specific stochastic model such as Markov chain (Gavirneni & Tayur, 2001) and autoregressive time series (Graves, 1999). In other words, any statistical assumption about customer demand is not required to compute the replenishment quantity. The replenishment quantity is a function of compensation factor (CF) that has an effect of increasing or decreasing the replenishment amount, and at each replenishment period, a cost-minimizing CF value is automatically chosen among the candidate set by using the retrospective action-reward learning.

The remainder of this paper is organized as follows. Section 2 introduces the basic concepts of action-reward learning and explains its application to nonstationary VMI situation. In Section 3, the situation reactive algorithm is presented in detail with some formal definitions. In Section 4, a simulation environment is explained and the results of the simulation based experiments are presented with discussions. Finally, conclusions are provided in Section 5.

2. Basic concepts

2.1. Action-reward learning

The integral component of the situation reactive approach is action-reward learning. Popular reinforcement learning methods, such as Q-learning and temporal difference learning (Mitchell, 1997; Sutton & Barto, 1998), were developed for Markov decision process with incomplete information on state transition that is used for determining the best action for each visited state. In inventory control case, state corresponds to the amount of inventory remaining before replenishment decision is made, and the state is influenced by customer demand pattern. However, due to the assumption of time-independent state change, the Markovian learning models cannot be directly applied to a supply chain where the behavior of customer demand is unstable over time. Although the action-reward learning is simpler than the Markovian learning models, it is capable of learning the best among several possible actions in a dynamically changing environment.

Typical learning model adopts single-update-at-a-time mechanism. That is, at period t, the model updates the performance measure of action selected at the previous period $t - 1$, because the reward of the action can be only discovered at t. The updates of the other actions are delayed until they are selected probabilistically in the future. In this learning mechanism, large number of actions critically deteriorates the reliability of their performance measures, which results in a poor control performance. In this research, we propose a multiple-update-at-a-time mechanism named retrospective analysis by extending the asynchronous learning mechanism (Kim, Kwon, & Baek, 2008) which was developed to control order interval between two supply chain members. However, due to irregular order interval, the characteristics of the supply chain considered in their model, the asynchronous learning mechanism has a limitation that only some of the actions that satisfy an update condition are eligible for update simultaneously. On the other hand, the retrospective analysis proposed in this paper is capable of updating the performance measures of all actions at each period very efficiently, thereby the best action for the current period can be found with little computational burden and greater reliability.

2.2. Mapping to nonstationary VMI situation

Fig. 1 summarizes the learning in nonstationary VMI situation. The supplier plays the role of an agent and its mission is to deliver the right amount of products (Q_t) in the beginning of each replenishment period to the retailer so that the long run average inventory cost is minimized. The replenishment quantity is determined based the forecasted mean demand during the next order period, a CF value, and the inventory amount of retailer remaining at the end of the current period. The CF value that should be chosen in the beginning of each period by the supplier is defined as an
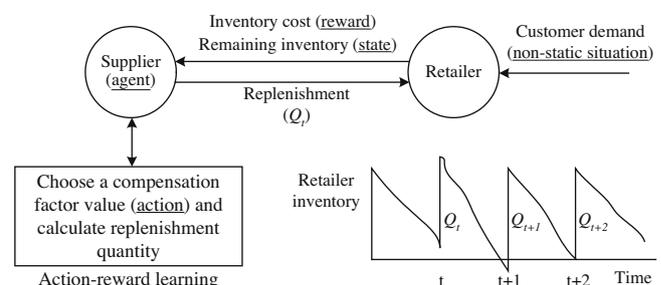

Fig. 1. Action-reward learning for nonstationary VMI model.