Expert Systems with Applications 36 (2009) 6520–6526

Contents lists available at ScienceDirect

Expert Systems with Applications

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

Case-based reinforcement learning for dynamic inventory control in a multi-agent supply-chain system

Chengzhi Jiang *, Zhaohan Sheng

Department of Management and Engineering, Nanjing University, 22 Hankou Road, Nanjing 210093, PR China

ARTICLE INFO

Keywords:
Inventory control
Reinforcement learning
Supply-chain management
Multi-agent simulation

ABSTRACT

Reinforcement learning (RL) appeals to many researchers in recent years because of its generality. It is an approach to machine intelligence that learns to achieve the given goal by trial-and-error iterations with its environment. This paper proposes a case-based reinforcement learning algorithm (CRL) for dynamic inventory control in a multi-agent supply-chain system. Traditional time-triggered and event-triggered ordering policies remain popular because they are easy to implement. But in the dynamic environment, the results of them may become inaccurate causing excessive inventory (cost) or shortage. Under the condition of nonstationary customer demand, the S value of (T, S) and (Q, S) inventory review method is learnt using the proposed algorithm for satisfying target service level, respectively. Multi-agent simulation of a simplified two-echelon supply chain, where proposed algorithm is implemented, is run for a few times. The results show the effectiveness of CRL in both review methods. We also consider a framework for general learning method based on proposed one, which may be helpful in all aspects of supply-chain management (SCM). Hence, it is suggested that well-designed "connections" are necessary to be built between CRL, multi-agent system (MAS) and SCM.

1. Introduction

Supply-chain management (SCM) has been providing competitive advantages for enterprises in the market. In that, inventory control plays an important role and has been attracting attentions from many researchers in recent years. Some known inventory control policies are studied and improved for all aspects, such as reduced cost, more flexibility. Chen, Li, Marc Kilgour, and Hipel (2006) introduce a case-based multi-criteria ABC analysis, that improves on this approach by accounting for additional criteria, such as lead time and criticality of SKUs. This procedure provides more flexibility to account for more factors in classifying SKUs. Lee and Wu (2006) propose the statistical process control (SPC) based replenishment method, in which inventory rules and demand rules are developed to determine the amount of order replenishment for solving order batching problem. This control system performs well at reducing backorders, and bullwhip effect. Yazgı Tütüncü, Aköz, Apaydın, and Petrovic (2007) present new models for continuous review inventory control in the presence of uncertainty. The optimal order quantity and the optimal reorder point are found to minimize the fuzzy cost.

On the other hand, different inventory management systems could be designed according to a specific industry or environment. Aronis, Magou, Dekker, and Tagaras (2004) apply Bayesian approach to forecasting the demand for spare parts of electronic equipment, providing a more accurate determination on stock level for satisfying negotiated customer service level. Ashayeri, Heuts, Lansdaal, and Strijbosch (2006) also develop cyclic production-inventory optimization models for the process manufacturing industry. ElHafsi (2007) shows that optimal inventory allocation policy in an assemble-to-order system is a multi-level state-dependent rationing policy. Díez, Erik Ydstie, Fjeld, and Lie (2008) design model-based controllers based on discretized population balance (PB) models for particular processes, which are encountered in almost any branch of process industries. Kopach, Balcıoğlu, and Carter (2008) revisit a queuing model and determine an optimal inventory control policy using level crossing techniques in blood industry.

Meanwhile, identifying factors affecting inventory management performance such as cost and demand also assists in designing the controllers. Andersson and Marklund (2000) introduce a modified cost structure at the warehouse, and then multi-level inventory control problem can be decomposed to single-level problems. By applying a simple coordination procedure to them, the near optimal solution is obtained. Zhang (2007) studies an inventory control problem under temporal demand heteroscedasticity, which is found to have a significant influence on firm's inventory costs. Chiang (2007) uses dynamic programming to determine the optimal control policy for a standing order system. Yazgı Tütüncü et al. (2007) make use of fuzzy set concepts to treat imprecision regarding the costs and probability theory to treat customer

* Corresponding author. Tel.: +86 13851833708.
E-mail address: tony_tiaotiao@hotmail.com (C. Jiang).

0957-4174/$ - see front matter © 2008 Elsevier Ltd. All rights reserved.
doi:10.1016/j.eswa.2008.07.036

demand uncertainty. Additionally, Maity and Maiti (2007) devise the optimal production and advertising policies for an inventory control system considering inflation and discounting in fuzzy environment.

It is observed that in recent researches mentioned, mathematical or analytical models are preferred, such as Bayesian approach (Aronis et al., 2004), Utility Function Method (Maity & Maiti, 2007), fuzzy set concepts (Yazgı Tütüncü et al., 2007) and Autoregressive and Integrated Moving Average and Generalized Autoregressive Conditional Heteroscedasticity (Zhang, 2007). This kind of method provides strict deduction, which usually involves complicated notations and equations under assumptions. However, on one hand, the problem may be time-varying under dynamic environment, especially in the evolving system like supply chain where the solution in one time may be not suitable for another time. On the other hand, those models are too difficult for managers to implement in the real enterprises because of the complicated calculations involved. This requires the learning ability to enrich one's experience continuously in order to make reasonable decisions. Reinforcement learning (RL) is an approach to machine intelligence that combines the fields of dynamic programming and supervised learning to yield powerful machine-learning systems (Harmon & Harmon, 1996). Chi, Ersoy, Moskowitz, and Ward (2007) demonstrate and validate the applicability and desirability of using machine learning techniques to model, understand, and optimize complex supply chains. To make the best use of learning methods, intelligent entities are the necessary carriers. Multi-agent Systems (MAS) seem to be a good choice where the agents are characterized of intelligence, autonomy, interactive and reactivity. Liang and Huang (2006) develop a multi-agent system to simulate a supply chain, where agents are coordinated to control inventory and minimize the total cost of a supply chain. Govindu and Chinnam (2007) propose a generic process-centered methodological framework for analysis and design of multi-agent supply chain systems.

Therefore, this paper proposes a reinforcement learning algorithm combined with case-base reasoning (CRL) in a multi-agent supply-chain system. Similar research is carried out by Kwon, Kim, Jun, and Lee (2007). They suggest a case-based myopic reinforcement learning algorithm for satisfying target service level using vendor managed inventory model. And in this paper, we are trying to provide a simpler learning method with similar or better performance, which could be used more widely and easier to implement by managers. Furthermore, the "connections" are strongly recommended to be built between CRL, MAS and SCM, thus a generic reinforcement learning method is also suggested.

The remainder of this paper is organized as follows. Section 2 explains the multi-agent supply-chain model including the inventory control problem. Section 3 presents the CRL algorithm in more detail. Simulation environment for measuring the performance of CRL is explained and the results are presented in Section 4. Section 5 considers a generic RL method based on the proposed one. Finally, the conclusion and future research are provided in Section 6.

2. Multi-agent supply-chain model

A simplified two-echelon supply chain consisting of multiple customers and retailers is designed for demonstration (see Fig. 1). It is assumed that each retailer receives all ordered stocks from suppliers in a fixed lead time regardless of the amount of order. And it faces nonstationary customer demand under two conditions:

(i) Each retailer has a fixed group of customers whose demand is nonstationary.
(ii) Each customer is free to choose one retailer in each period, i.e., in a competitive market.

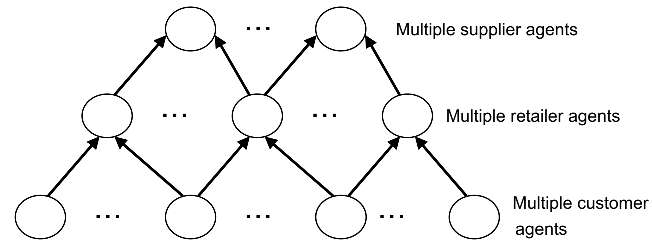

Fig. 1. The supply-chain model under consideration.

The second condition requests additional customer selecting standards and strategies for retailers trying to attract customers or maximize their profits. The former one adopts simplified motivation function proposed by Zhang and Zhang (2007)

$$M_i = PS_i \times P_i + QS_i \times Q_i + ft_i \times in_i. \tag{1}$$

M_i motivation of retailer i ($i = 0$ to $N_{\mathrm{retailer}} - 1$) of a customer agent. PS_i presents customer agent's price sensitivity parameter to retailer i, while QS_i is the quality sensitivity parameter and ft_i is the follower tendency. P_i and Q_i are the price and quality of retailer i respectively. in_i is the influence received from some other customer agents as friends with the respect to retailer i. Eq. (1) is further specified as follows:

$$M_i = (-\alpha^{P_i - P_{\mathrm{ave}}} + k) \times P_i + \left(\beta^{|Q_i - Q_{\mathrm{ave}}|} + L\right) \times Q_i + ft \times in_i, \tag{2}$$

where, $\alpha > 1$, $0 < \beta < 1$, P_{ave} and Q_{ave} are the average price and quality provided by retailers. k is a constant presenting the price sensitivity which is varied according to the social-status of customers. And L is the corresponding quality sensitivity constant of customers. The calculation of in_i is treated as: each customer has a list of influence out from positive to negative corresponding to its own rank of retailers. in_i equals the added value of influence out of retailer i from friend agents.

And a simple adjustment strategy for retailer i is used here as follows:

if its market share is below average, then $P_i = P_i - p$;
else if its market share is above average, then $P_i = P_i + p$;
else no change is made.

Under both conditions, the demand that is not met immediately is treated as lost sales. However, under condition 2, for each customer has its own rank of retailers according to motivation function values, it may select the next one when the retailer of higher rank has insufficient inventory.

Besides, all retailers firstly use periodical review order-up-to (T, S) system and then order-quantity reorder point (Q, S) system for inventory management. And the order-up-to level and reorder point in both systems are learned, respectively using CRL for satisfying target service level (see Figs. 2 and 3). The goal of each retailer is to satisfy its target service level predefined while trying to cut excessive inventory. In this paper, the fill-rate type service level is adopted.

It can be seen that the CRL in (T, S) system takes place before lead time, because it will learn the rewards before and suggest

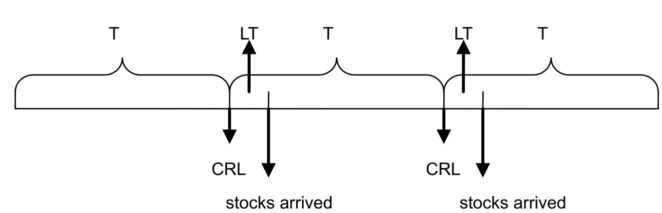

Fig. 2. (T, S) inventory replenishment mechanism.