Computers & Operations Research 36 (2009) 1703–1710

Contents lists available at ScienceDirect

Computers & Operations Research

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

Differentiated service inventory optimization using nested partitions and MOCBA

Ek Peng Chew*, Loo Hay Lee, Suyan Teng, Choon Hwee Koh

Department of Industrial and Systems Engineering, National University of Singapore, 10 Kent Ridge Crescent, Singapore 119260, Singapore

ARTICLE INFO

Available online 29 April 2008

Keywords:
Nested partitions method
Multi-objective computing budget allocation
Pareto optimality
Differentiated service
Inventory

ABSTRACT

In this paper, we consider a differentiated service inventory problem with multiple demand classes. Given that the demand from each class is stochastic, we apply a continuous review policy with dynamic threshold curves to provide differentiated services to the demand classes in order to optimize both the cost and the service level. The difficult features associated with the problem are the huge search space, the multi-objective problem which requires finding a non-dominated set of solutions and the accuracy in estimating the parameters. To address the above issues, we propose an approach that uses simulation to estimate the performance, nested partitions (NP) method to search for promising solutions, and multi-objective optimal computing budget allocation (MOCBA) algorithm to identify the non-dominated solutions and to efficiently allocate the simulation budget. Some computational experiments are carried out to test the effectiveness and performance of the proposed solution framework.

1. Introduction

The inventory systems with several demand classes can be found in many applications. Traditionally, all customer demands are treated equally and served on a first-come-first-serve basis. In a differentiated service inventory system, customers are classified into different groups according to their importance to the decision makers. For example, some customers want to have low stock-out, and they are willing to pay a higher price for it. Obviously, these customers are more valuable and therefore they should have higher priority.

The differentiated service inventory problem is to determine how to replenish inventories and how to allocate these inventories to different demand classes according to some performance measures such that each demand class is offered with different service level. Several decision policies aiming at differentiating demand classes and offering different services have been proposed in the literature. The multi-class inventory problems were first analyzed by Veinott [1] under a periodic review policy, where at every period the inventory is used to satisfy the demand of different demand classes in the order of the highest priority customer to the lowest priority customer. Another policy that has been proposed is called the critical level policy. In this policy with m demand classes, there are $m-1$ critical levels. When demand from a particular class arrives, it will be satisfied if the current inventory level is higher than the critical level associated with that demand class; otherwise it will be rejected. The demand class with the highest priority has a zero critical level. For example, Dekker et al. [2] applied the critical level policy under the $(S-1,S)$ inventory system with general lead time and lost sales, and Deshpande et al. [3] considered a threshold clearing mechanism for two demand classes under the (s,Q) inventory policy.

A dynamic threshold policy where the critical level depends on the remaining time before the next replenishment arrives is also proposed. Some of these works can be found in [4–7]. In these policies, the decision of accepting or rejecting an incoming demand is made based on the customer class, the inventory on hand, and the remaining time before the next replenishment arrives. Chew et al. [7] show that the threshold level is non-decreasing as the remaining time before the next replenishment arrives becomes less. Their approach is similar to airline revenue management, where the number of seats to be sold is controlled based on both the remaining time before departure and the type of customers.

In this study, we consider a differentiated service inventory problem with m different demand classes. We assume that the demand for each class is stochastic, and the inventory is replenished according to a continuous (s,Q) inventory model. Under the dynamic threshold policy proposed in [7], the problem is to obtain a set of non-dominated reorder point s and order quantity Q with the best combination of cost and service level. Here the cost is defined in terms of average annual cost, which consists of setup cost, inventory holding cost and backorder cost. The service level is defined in terms of average backorder of each customer class, which is calculated as the average number of backorders per year. The paper is organized as follows. In Section 2, we propose an integrated solution framework which can find a non-dominated set of inventory

* Corresponding author. Tel.: +65 6874 6554; fax: +65 6777 1434.
E-mail address: isecep@nus.edu.sg (E.P. Chew).

0305-0548/$ - see front matter © 2008 Elsevier Ltd. All rights reserved.
doi:10.1016/j.cor.2008.04.006

(s, Q) policies to the differentiated service inventory problem. Section 3 presents some computational results and analysis, and finally conclusions and future research directions are summarized in Section 4.

2. Solution framework for the differentiated service inventory problem

The differentiated service inventory problem has some difficult features: large search space, multiple objectives and variability involved in the objectives. In this section, we present an integrated solution framework for solving this problem. The solution framework uses simulation to estimate the performance measures, nested partitions (NP) method [8] to explore the solution space for more promising solutions, and a statistical selection procedure to efficiently allocate simulation budget and identify the non-dominated Pareto set of solutions.

2.1. Overview of the integrated solution framework

2.1.1. A discrete-event simulation model

The differentiated service inventory problem involves two performance measures, namely the average annual cost and the expected backorder. Under the dynamic threshold policy with m customer classes, it is difficult to derive an explicit analytical model for the (s, Q) inventory policy which can simultaneously find the non-dominated set for the two objectives. Therefore, we resort to simulation models to evaluate the performance measures of a given (s, Q) inventory policy.

We build a discrete-event simulation model based on the following assumptions:

- Lead time for the order arrival is constant.
- All the unfulfilled customer demands are backordered.
- Customer arrivals follow a Poisson distribution.

Given a reorder point s and order quantity Q, the simulation model simulates the customer arrival events, the ordering events, and the replenishment order arrival events. Moreover, each time when a customer demand of a certain class is generated in the simulation model, the threshold value for the respective class is computed according to [7]. If the threshold value for the respective class is below the current inventory level, then the demand is satisfied; otherwise it is rejected and considered as a backorder. The output of the simulation model is the average annual cost and the average backorder for each customer class. Hence, we can simulate different (s, Q) values and select the set of non-dominated ones by comparing the performance measures (average annual cost and average backorder).

2.1.2. The NP search method

Simulation can only be used to evaluate the performances of given design alternatives, i.e., it lacks the ability of exploring more promising solutions in the search space. This problem is trivial when the search space is small, as exhaustive search is possible. In our problem, however, the search space can be very huge, considering we have m customer classes and have to search the whole feasible region of both reorder point s and the order quantity Q.

NP method [8] is a relatively recently developed algorithm for global optimization. Though NP is initially developed for optimization problems with deterministic objective function, it has been successfully integrated with statistical selection techniques to solve single objective simulation optimization problems [9,10]. In this study, our problem is to search optimal (s, Q) values within the feasible region which is an area formed by upper bounds of both s and Q. It is quite convenient and natural to employ NP to partition the search space iteratively and focus on the more promising region in searching for better solutions. Thus, we employ NP as a main search engine in this study.

2.1.3. A statistical selection procedure

Since we use simulation to evaluate the performance measures, variability involved in the simulation output may pose problems when we apply the NP method. For instance, in NP, promising index for each region is calculated based on performance measures of the samples picked in that region, and future promising region is selected according to the promising indices calculated. With inaccurate estimation of the performance measures of the samples, the promising region that future search focuses on may turn out to be less promising, leading NP to non-productive regions. To address this issue, a statistical selection procedure is required to determine: (1) the right number of simulation replications to run for each solution so that the performance measures can be more accurately estimated without wasting simulation budget; (2) how to guarantee the best set of solutions selected are the true best with high confidence. There are a number of statistical selection procedures for the single objective problem [11–16]. In this study, as we are considering a multi-objective problem which requires a Pareto set of non-dominated solutions, we employ the multi-objective optimal computing budget allocation (MOCBA) procedure presented in Lee et al. [17] for the above purposes. A brief description of the MOCBA procedure is given in the following section.

2.2. The MOCBA algorithm

Given a set of n designs each with H performance measures which are evaluated through simulation, the MOCBA algorithm developed in Lee et al. [17] is to efficiently allocate simulation replications to the designs so that the non-dominated Pareto set of solutions can be found with high confidence at the least expense of simulation budget.

Assume that design i is evaluated in terms of H performance measures $\tilde{\mu}_{ik} (k = 1, 2, \ldots, H)$ through simulation. Within a Bayesian framework, $\tilde{\mu}_{ik}$ is considered as a random variable following posterior distribution derived from simulation output. We use the following performance index to measure how non-dominated design i is

$$\psi_i = \prod_{j=1, j\neq i}^{n} \left[1 - \prod_{k=1}^{H} P(\tilde{\mu}_{jk} \leqslant \tilde{\mu}_{ik}) \right]. \tag{1}$$

In MOCBA, the quality of the Pareto set depends on whether designs in Pareto set (S_p) are all non-dominated and designs in non-Pareto set ($\bar{S}_p$) are all dominated. We evaluate it by Type I error (e_1) and Type II error (e_2) which can be bounded as follows:

$$e_1 \leqslant ae_1 = \sum_{i \in \bar{S}_p} \min_{\substack{j \in S \\ j \neq i}} \left[1 - \prod_{k=1}^{H} P(\tilde{\mu}_{jk} \leqslant \tilde{\mu}_{ik}) \right], \tag{2}$$

$$e_2 \leqslant ae_2 = \sum_{i \in S_p} \left(1 - \prod_{\substack{j \in S \\ j \neq i}} \left[1 - \prod_{k=1}^{H} P(\tilde{\mu}_{jk} \leqslant \tilde{\mu}_{ik}) \right] \right). \tag{3}$$

To get the true Pareto set with high probability, we need to minimize both Type I and Type II errors. In the MOCBA algorithm of Lee et al. [17], this is done by iteratively allocating the simulation replications until some stopping criteria are met according to the asymptotic allocation rules stated in the following Lemmas.