J. Korea Saf. Manag. Sci. Vol. 16 No. 4 December 2014
http://dx.doi.org/10.12812/ksms.2014.16.4.323

ISSN 1229-6783(Print)
ISSN 2288-1484(Online)

강화학습 기반의 다단계 공급망 분배계획

권 익 현*
*인제대학교 산업경영공학과

Reinforcement leaning based multi-echelon supply chain distribution planning

Ick-Hyun Kwon*
*Department of Industrial and Management Engineering, Inje University

Abstract

Various inventory control theories have tried to modelling and analyzing supply chains by using quantitative methods and characterization of optimal control policies. However, despite of various efforts in this research filed, the existing models cannot afford to be applied to the realistic problems. The most unrealistic assumption for these models is customer demand. Most of previous researches assume that the customer demand is stationary with a known distribution, whereas, in reality, the customer demand is not known a priori and changes over time. In this paper, we propose a reinforcement learning based adaptive echelon base-stock inventory control policy for a multi-stage, serial supply chain with non-stationary customer demand under the service level constraint. Using various simulation experiments, we prove that the proposed inventory control policy can meet the target service level quite well under various experimental environments.

Keywords : Multi-echelon, Serial Supply Chain, Base-stock, Non-stationary Demand, Service Level

1. 서 론

공급망의 재고관리는 기업의 비용 절감 측면에서 오랫동안 중요한 문제로 연구되어 왔으며, 기업들은 적정량의 재고수준을 유지함으로써 고객의 수요에 빠르게 대처할 수 있을 뿐만 아니라 나아가 고객 만족을 통한 기업 이미지 상승을 도모하여 왔다. 그러나 재고과잉은 추가적인 재고유지 비용의 발생을 초래하고, 재고부족은 고객 서비스 수준의 하락 및 판매기회 상실비용의 발생을 초래하므로 적정한 재고수준을 유지할 수 있는 효율적인 재고관리에 대한 관심은 점점 높아지고 있다[9,12].

최근 들어 인터넷의 광범위한 보급은 전자상거래의 활성화를 초래하였고, 전자상거래를 이용하는 온라인상에서의 고객수요는 작은 가격차이만으로도 쉽게 변화하기 때문에 전반적으로 고객수요가 매우 불안정한 형태로 나타나게 되었다. 또한, 비슷한 품질과 기능을 가진 다양한 제품의 출시로 인한 제품 차별성 부재는 고객수요의 불안정성을 증폭시키는 원인이 되었다.

†이 논문은 2012년 정부(교육과학기술부)의 재원으로 한국연구재단의 지원을 받아 수행된 연구임 (NRF-2012S1A5A8024848).
†Corresponding Author : Ick-Hyun Kwon, Department of Industrial and Management Engineering, Inje University, 197 Inje-Ro, Gimhae-Si, Gyeongsangnam-Do, 621-749 Tel: 055-320-3992, E-mail: ikwon@inje.ac.kr
Received July 20, 2014; Revision Received December 19, 2014; Accepted December 22, 2014.

고객수요의 불안정성은 수요예측의 불확실성을 증가시키는 원인이 되고, 수요예측의 불확실성 증가는 재고부족이나 불필요한 재고과잉을 유발시키며 이로 인한 재고비용의 발생을 초래한다. 따라서 많은 기업들이 불필요한 재고비용을 줄이기 위한 연구에 관심을 집중시키고 있으며, 특히 불안정한 고객수요 환경 하에서 재고비용 감축 및 고객 서비스 수준 충족을 위한 연구에 많은 노력을 기울이고 있다.

지금까지 공급망의 재고관리에 관한 대부분의 연구는 안정적인(stationary) 고객수요 하에서 정확한 분포의 형태 등이 사전에 알려져 있다고 가정한 반면, 현실에서의 고객수요는 그 분포를 미리 알 수 없으며 시간의 경과에 따라 예측할 수 없이 변화하는 특성을 갖는 것이 일반적이다. 또한 기존 대부분의 연구는 안정적인 수요를 따르는 비교적 단순한 공급망을 대상으로 확률적 모델이나 분석적 모델 등에 기반하는 수리적인 접근방법을 통해 최적해 또는 근사 최적해를 도출하고자 하였다. 이러한 기존의 공급망 관련 연구 방법들은 비교적 정확한 결과값을 제시하지만, 공급망의 형태가 복잡해지거나 다른 제약조건이 추가될 경우 제한된 시간에 결과를 도출하기에는 한계가 있어 비교적 제한된 범위에 적용될 수밖에 없으며 그 결과를 실제 현장에 적용하기에 많은 어려움이 따르는 한계점이 존재한다[8].

본 논문에서는 이러한 기존 연구의 한계점을 극복하고 보다 실제적인 공급망에 적용 가능한 재고정책을 위해 불안정한 고객 수요가 발생하는 공급망을 대상으로 하는 적응형 재고정책을 제안하고자 한다. 본 연구에서는 기초재고 정책(base-stock policy)을 사용하는 모델을 제안하며, 이를 통해 주어진 목표 서비스 수준 제약을 만족시키는 것을 목적으로 한다. 대상이 되는 공급망의 형태는 연속형 공급망(serial supply chain)이다. 본 연구에서 제안하는 적응형 재고통제 정책은 기존의 수요예측 모델(e.g., exponential smoothing, moving average 등)을 이용하여 시간에 따라 변화하는 고객 수요를 효과적으로 예측하고, 예측된 고객 수요를 실시간으로 활용하여 목표 서비스 수준을 준수하기 위한 분배량 및 분배 시점 등을 동적으로 조정하는 절차를 따른다.

공급망을 구성하는 각 객체들은 공급망의 다른 구성단위들과 서로 유기적으로 연관되어서 서로 영향을 주고받으며 상위 구성단위는 하위 구성단위의 계획 수립에 제약을 주는 특성 때문에, 한 구성단위 차원의 관점에 국한되어서는 전체 공급망의 최적화를 가져올 수 없다. 본 연구에서는 강화학습(reinforcement learning) 기법[10]의 일종인 행동-보상 학습(action-reward learning)을 이용하여 시간에 따라 변화하는 고객 수요에 대처할 수 있도록 재고통제 모수인 기초재고 수준(base-stock level)을 적응적으로 조정하는 방법을 적용하고자 한다. 행동-보상 학습이란 에이전트(agent)의 시행착오(trial-and-error)를 이용한 인공 지능형 학습 방법으로써 주어진 환경 하에서 가장 높은 보상을 얻을 수 있는 행동이 다음 계획기간에 선택될 확률이 높아지도록 학습함으로써, 고객 수요의 변화에 대처하여 최적의 행동이 선택될 수 있도록 지원한다.

안정적인 수요 하에서의 서비스 충족문제는 이미 상당한 연구결과가 축적되어 있으며 다양한 선행연구를 통해 체계적으로 정리되어 발표된바 있다[2]. 불안정한 수요에 대한 공급망관리에 대한 연구는 안정적인 수요에 비하여 상대적으로 많지 않는 실정인데, 대표적인 연구로 Gavirneni and Tayur[4]는 고객 수요가 불안정한 형태를 나타내는 원인을 다양한 범주(예: 경제상황의 변화, 계절적인 요인, 제품의 수명주기 단축 등)로 구분하고 각각의 범주에 해당되는 기존 연구들을 체계적으로 분류하고 자세히 설명하였다. 불안정한 고객수요하의 공급망관리에 관한 연구로, Kim et al.[6]은 2단계 연속형 공급망(two-stage serial supply chain)을 대상으로 계획 기간 동안 전체 공급망에서 발생하는 재고비용의 총합을 최소화하기 위한 2가지 모델(centralized and decentralized models)을 제안한 바 있다. 이 논문에서는 불안정한 고객수요의 변화에 대처하기 위해 강화학습을 사용하여 공급자(supplier)의 안전 리드타임과 소매점(retailer)의 안전재고를 적응적으로 조절하는 방법을 제안하고, 제안된 2가지 모델의 장단점을 비교, 분석한 바 있다. Kwon et al.[8]은 CBR(case-based reasoning)과 강화학습 기법을 이용하여 생산용량의 제한이 있는 2단계 depotless 형태의 분배시스템을 대상으로 비용에 대한 고려 없이 각 노드들의 주어진 서비스 수준을 준수할 수 있도록 하는 적응형 재고정책을 제안한 바 있다. 본 연구와 유사한 최근의 연구로 Jiang and Sheng[5]은 목표 서비스 수준 만족을 목적으로 하는 2단계 분배형 공급망(two-stage distribution-type supply chain)을 복수 에이전트(multi-agent)를 통해 모델링하고 CBR(case-based reasoning)과 강화학습을 이용하는 방법론을 제안하였다. 그러나 이들은 리드타임을 고려하지 않아 제안된 공급망 모형이 충분한 현실성을 반영하는지를 평가하기에는 부족한 면이 많은 것으로 판단된다.

2. 본론

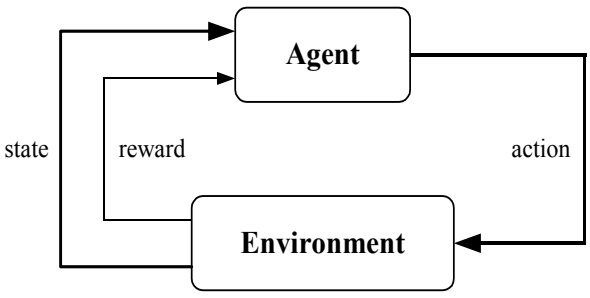
본 연구에서는 불안정한 고객 수요가 발생하는 다단계 공급망에 대한 적응형 재고통제 정책을 제안하고자 한다. 본 연구에서는 기초재고 정책(base-stock policy)을 사용하는 모델을 제안하며 이를 토대로 주어진 목표 서비스 수준 제약을 만족하는 분배계획을 수립하는 것을 목표로 한다.

본 논문에서 제안하는 연구와 대부분의 기존 연구와의 가장 뚜렷한 차이점은 고객수요의 형태에 대한 가정에 있다. 본 연구에서 가정하는 고객 수요는 현실에서 발생 가능한 동적인 형태의 불안정한 고객 수요를 의미한다. 이러한 불안정한 형태의 고객 수요는 기존 논문에서 주로 고려하는 ARIMA(autoregressive integrated moving average)나 moving average process 등과 같은 시계열 모형이나 통계적 분포가 아니며, 시간에 따라서 그 평균과 분산 등이 변화하는 것이다. 또한 이러한 고객수요의 변화 행태는 사전에 알지 못하는 것으로 가정한다. 이러한 불안정한 고객 수요는 서비스 수준의 준수와 재고비용 감소를 더욱 어렵게 만드는 요소가 된다. 이러한 이유로 인하여 종전의 확정적(deterministic)이거나 변하지 않는 수요분포에 기초하여 재주문점, 안전재고 등을 정한다는 개념은 실제적인 공급망 계획에 있어서 사용이 부적절하고 새로운 서비스 수준의 충족 방법이 나와야 하며 이를 기초로 재고계획이 수립되어져야 한다. 본 연구는 실 상황의 동적이며 불안정한 특성을 갖는 고객수요의 행태하에서 주어진 서비스 수준을 보장하는 재고정책을 다룬다는 점에서 타 연구들과 차별되는 독특함을 보인다. 본 연구에서 제안하고자 하는 재고통제 기법은 수요변화에 따라 재고관리를 위해 필요한 모수를 적응적으로 제어하는 특징이 있다. 이를 위하여 기계학습 방법론의 일종인 강화학습(reinforcement learning) 기법을 적용한다. 본 연구에서는 연속형 공급망(serial supply chain)의 최하위 단계인 소매점에서부터 최상위 노드의 순으로 노드 번호를 부여하기로 한다. 물류는 외부의 무한 용량을 갖는 공급자로부터 제품을 공급 받는 최상위 노드로 부터 최하위 노드까지로 이동하며, 최하위 노드에서만 고객의 수요가 발생한다. 이들은 고정되어 있으며, 단위 계획기간의 정수배 형태로써 각기 다른 시간을 가질 수 있다.

2.1 접근방법

공급망에서의 효율적인 재고관리 방법을 제안한 기존의 대부분의 연구들은 재고관리를 위한 수리적 모형을 정의하고, 이를 기반으로 최적해를 도출하는 방법을 제시하였다[1,3,13]. 그러나 수리적 모형을 이용한 방법은 고객수요가 안정적인 분포를 따른다는 가정 하에 최적해를 도출하는 방법을 제시하고 있기 때문에 고객수요가 불안정할 경우에는 적용하기 힘들다는 문제점을 지닌다. 본 연구에서는 이러한 수리적인 접근 방법의 단점을 보완하면서, 보다 다양한 형태의 공급망 네트워크에 대해 전체 재고 비용을 최소화시키면서 목표 서비스 수준을 만족하기 위해 강화학습의 일종인 행동-보상 학습(action-reward learning)을 통해 제어하는 방법론을 제안하고자 한다.

일반적으로 행동-보상 학습은 미리 정의된 상태별로 최적의 행동을 결정하는 Markov decision process (MDP)를 위한 기계학습(machine learning) 방법론이다. 이러한 MDP를 풀기 위한 방법으로서 강화학습은 기계학습과 인공지능 연구자들에 의해서 90년대 들어와서 각광을 받기 시작한 연구 분야이다. <Figure 1>에서 보는바와 같이 행동-보상 학습(action-reward learning)은 여러 가지 가능한 행동(action)들 중에서 최적의 행동을 선택하기 위해 시행착오(trial and error)를 통한 반복학습을 기본 원리로 한다. 행동-보상 학습에서는 에이전트(agent)라 불리는 학습자(learner) 또는 의사결정자(decision maker)가 불확실한(non-deterministic) 환경에서 특정한 행동을 선택하면 그에 따라 대상 환경의 상태(state)가 변화하고, 상태의 변화에 따라 해당 행동에 대한 보상(reward)이 결정된다[10].



[Figure 1] Principle of action-reward learning

따라서 에이전트는 행동에 따른 보상의 합을 최대화하기 위해 가능한 행동들 중에서 보상 값이 가장 높은

행동을 선택하고, 선택된 행동에 의해 변화된 상태에 따라 보상 값이 수정되는 과정을 반복함으로써 에이전트는 불확실한 환경에서의 학습을 수행할 수 있게 된다. 본 연구에 행동-보상 학습을 적용하면, 공급망의 의사결정자(decision maker)가 에이전트가 되고 불안정한 고객수요를 갖는 공급망이 환경이 된다. 또한 주문주기마다 선택되어야할 보상계수는 행동으로 정의될 수 있으며, 선택된 행동에 의해 결정되는 다음 주문시점에서의 재고수준(재고부족 또는 재고과잉)이 상태로 정의될 수 있다. 그리고 재고수준에 따라 발생하는 서비스 수준이 행동에 따른 보상으로 정의될 수 있다.

2.2 제안된 방법론

본 논문에서는 수요 변화에 따라 각 노드의 계층 기초재고 수준(echelon base stock level)을 동적으로 계산하는 방법을 적용한다. 본 연구에서 사용하는 기호는 다음과 같다.

- β : 목표 서비스 수준
- L_i : 노드 i의 리드타임
- $L_{[i,j]}$: 노드 i부터 노드 j까지의 누적 리드타임, $L_{[i,j]} = \sum_{k=i}^{j} L_k$
- D_t : $[t,\ t+1)$ 기간 동안의 실제 고객 수요
- $\widehat{D_{t,t'}}$: t기간에 예측된 t'기간의 예측 수요
- $\widehat{\sigma_{i,t}}$: $[t-L_{[1,i]},\ t-1]$ 동안의 고객 수요에 대한 예측 오류의 표준편차
- $IP_{i,t}$: 노드 i의 t기간 초의 계층 재고상태(echelon inventory position)
- $S_{i,t}$: 노드 i의 t기간에서의 계층 기초재고 수준(echelon base-stock level)
- $sd_{i,t}$: 노드 i의 t기간에서의 안전공급량(safety delivery quantity)
- ρ_k : 보상계수(compensation factor; CF) 값
- Θ : CF의 집합($\Theta = \{\rho_1, \rho_2, \cdots, \rho_K\}$)

$$S_{i,t} = \sum_{k=1}^{L_{[1,i]}} \widehat{D_{t,t+k}} + sd_{i,t},\ sd_{i,t} = \rho_k \cdot \widehat{\sigma_{i,t}} \quad (1)$$

$$Q_{i,t} = S_{i,t} - IP_{i,t},\ t = 1,2,3,... \quad (2)$$

$$IP_{i,t} = IP_{i,t-1} + Q_{i,t-1} - D_{t-1}, t = 1,2,3,... \quad (3)$$

수식 (1)에서와 같이 노드 i의 계층 기초재고 수준(echelon base-stock level)은 노드 1부터 노드 i까지의 누적 리드타임 동안의 예측수요와 안전공급량의 합으로 이루어진다. 계층 재고 정책(echelon stock policy)은 각 노드의 계층 재고(echelon stock)에 기반을 두어 주문량을 결정하는 주문 정책이다. 계층 재고란 해당 노드의 보유 재고와 이 노드를 통해 물품을 공급받는 모든 하위 노드들에 대한 보유 재고와 수송중(in-transit) 재고의 총합에서 최하위 노드의 재고이월을 뺀 수치로 정의된다. 이와 함께, 계층 재고 상태(echelon inventory position)는 해당 노드의 계층 재고에서 수송 중(in-transit) 재고를 더한 값을 나타낸다. 이와 같은 계층 재고 정책은 하위 단계의 재고량의 합에 기반 하는 재고 정책이다. 즉, 상위 단계의 주문량을 결정하는데 있어 하위 단계가 재고를 많이 가지고 있는 경우 상위 단계의 주문량을 낮춰주며 반대로 하위 단계의 재고가 적을 경우 상위 단계의 주문량을 높임으로써 공급망 상의 모든 단계의 재고량을 효과적으로 유지하게 한다[1].

수식 (2)에서와 같이 안전공급량(safety delivery quantity)은 다시 보상계수(compensation factor; 이하 CF) 값($\rho_k \in \Theta$)과 고객 수요에 대한 예측 오류의 표준편차($\widehat{\sigma_{i,t}}$)의 곱으로 이루어지며, 이는 누적 리드타임 동안의 고객수요의 변화를 반영하기 위한 목적으로 일반적으로 많이 사용된다[12]. 따라서 본 연구에서는 고객 수요의 변화에 따르는 적절한 CF를 선택함으로써 기초 재고 수준(base-stock level)을 동적으로 조정하는 방식을 적용하고자 한다. CF의 집합인 Θ은 0보다 작은 값에서부터 0보다 큰 값을 포함하는 방식으로 나타낼 수 있다(예: {-2.0, -1.9, ..., 1.9, 2.0}). 만약 지난 기간에 선택된 CF 값으로 인해 목표 서비스 수준을 달성하지 못하였을 경우, 현 시점에서는 보다 큰 CF 값을 선택함으로써 이러한 문제점을 해결할 수 있다.

고객 수요가 안정적인 정규분포를 따른다는 가정 하에서는 특정 서비스 수준을 만족시킬 수 있는 CF 값을 쉽게 계산할 수 있으며[1,11], 이 값은 시간이 흐름에 따라 변화지 않는다. 그러나 본 연구의 대상이 되는 공급망에서는 고객 수요가 불안정한 형태를 나타내기 때문에 기존의 방법을 통해 CF 값을 계산하기 어렵다. 즉, 고객 수요의 평균이나 표준편차가 시간에 따라 변하기 때문에 하나의 최적 CF 값이 존재하지 않으며, CF 값은 고객 수요의 변화에 맞추어 적응적으로 선택

되어져야 한다.

본 논문에서는 이와 같은 문제점을 해결하기 위한 적응형 재고관리 모델을 제시하고자 한다. 본 연구에서 제시하는 공급망 계획의 기본 방향은 기존에 개발된 예측 모델을 이용하여 시간에 따라 변화하는 고객 수요를 예측하고, 예측 수요를 기반으로 목표 서비스 수준을 준수하기 위한 각 노드별 계층 기초재고 수준(echelon base-stock level)을 결정하는 방법을 따른다. 이를 위해서는 시간에 따라 변화하는 고객 수요를 기반으로 CF 값이 동적으로 조정되어야 하는데, 본 연구에서는 이러한 문제를 해결하기 위하여 강화학습 기법의 일종인 행동-보상 학습(action-reward learning)을 이용하여 CF 값을 적응적으로 조정하는 방법을 제시하고자 한다.

행동-보상 학습은 여러 가지 가능한 행동(action)들 중에서 최적의 행동을 선택하기 위해 고안된 학습방법이다. 일반적으로 모든 시스템은 에이전트와 그의 환경으로 구성되어 있다고 볼 수 있다. 에이전트는 그가 처한 환경에서 자신의 목적을 달성하기 위한 최적의 의사 결정을 하여야 하는데 이를 위하여 시행착오(trial-error)를 반복적으로 시행하여 가장 좋은 정책을 찾게 된다.

본 연구에서 사용한 행동-보상 학습은 Sutton and Barto[10]를 참조하였으며 개념적인 수식은 다음과 같다.

$$\text{NewEstimate}(\rho_k) \leftarrow \text{OldEstimate}(\rho_k) + \text{StepSize}[\text{CurrentReward}(\rho_k) - \text{OldEstimate}(\rho_k)] \quad (4)$$

위 수식은 특정 행동 ρ_k에 대한 예상 보상값(NewEstimate(ρ_k))을 학습하는 식이다. 즉, 선택할 수 있는 행동의 수가 복수일 때, 현재 시점에서 선택된 행동에 대해서 얻어진 보상값(CurrentReward(ρ_k))을 이용하여, 선택된 행동에 대한 기존의 예상 보상값(OldEstimate(ρ_k))과의 차이의 일정비율(StepSize)만큼을 더해나가며 조정해 나간다.

위의 수식 (4)에 의해서 계산된 행동별 예상 보상값은 수식 (5)와 같은 softmax rule에 의해서 다음 시점에 선택할 새로운 행동을 확률적으로 선택하는 데에 사용된다. 보상 최대화의 경우 NewEstimate(ρ_k) 값이 증가할수록 수식 (5)의 오른쪽 분자값은 증가하기 때문에 ρ_k의 선택 확률은 높아지게 된다.

$$\Pr(\text{new action} = \rho_k) = \frac{e^{-1/NewEstimate(\rho_k)}}{\sum_{\rho_j \in \Theta} e^{-1/NewEstimate(\rho_j)}} \quad (5)$$

본 연구에서는 매 기간, 각 노드별 기초재고 수준(또는 CF 값)을 동적으로 결정해야 하는데, 이 경우 평균 서비스 수준이 CF 값을 선택하는 기준이 된다. 예를 들어 만약 $t-L_i$ 기간에서 노드 i에 대한 CF 값으로 ρ_k가 선택되었고, t기간의 실제 서비스 수준이 $\beta_{i,t}(\rho_k)$일 경우 (노드 i, t기간)의 ρ_k에 대한 예상 서비스 수준은 각각 아래와 같이 갱신된다.

$$\hat{\beta}_{i,t}(\rho_k) = \hat{\beta}_{i,t-1}(\rho_k) + StepSize[\beta_{i,t}(\rho_k) - \hat{\beta}_{i,t-1}(\rho_k)] \quad (6)$$

행동(본 연구에서는 CF 값)에 대해서는, 불안정한 고객 요구량에 적절히 적응하기 위하여 미리 설정된 범위내의 고정값을 대상으로 학습을 한다. 또한 매 시점마다 선택된 CF 값 이외의 다른 CF 값에 대한 결과값(서비스 수준)은 시간을 되돌려서(1기간 이전 시점) 이미 발생한 고객 수요를 이용하여 평가하는 retrospective analysis[6,7]를 적용함으로써 학습 효율을 향상시킬 수 있다. 즉, 매 기간 모든 CF 값에 대한 예상 서비스 수준을 동적으로 갱신해 나간다. 노드 i의 t기간에 대한 새로운 CF 값을 선택하기 위해서는 수식 (6)을 이용한다. 주어진 서비스 수준을 준수하면서 재고비용을 최소화하기 위해서는 $\widehat{\beta_{i,t}}$이 목표 서비스 수준에 가장 근사한 CF 값들 중에서 선택하도록 한다.

3. 실험 및 결과분석

본 논문에서 제안하는 모델의 성능 테스트를 위해 시뮬레이션을 수행하였다. 시뮬레이션의 대상이 되는 공급망은 4개의 노드로 구성된 4단계 연속형 공급망(4-stage serial supply chain)이다. 한 시행의 시뮬레이션은 2,000시간 단위만큼 시행되었고, warm-up 기간으로 100시간 단위를 사용하였다. 실험은 고객 수요의 불안정도(non-stationarity)와 리드타임에 따라서 설계하고 10회씩 시행하여 그 평균을 비교하였다.

본 논문에서는 기존의 불안정한 수요를 고려한 논문[5,6,8]에서 일반적으로 활용하는 실험조건을 참조하여 모든 수요 분포는 정규분포 $N(m, s^2)$을 따른다고 가정하였으며 수요의 평균(m)과 표준편차(s)의 변경 간격과 범위에 따라서 아래와 같이 평균의 변화도

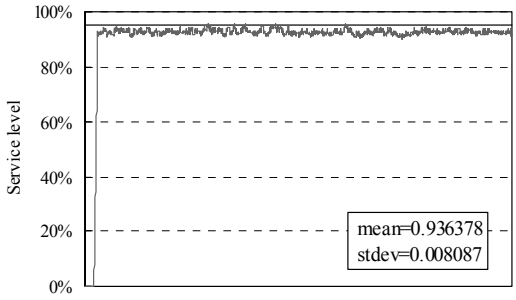
(mean variation: MV)가 세 가지로 나누어지도록 설계하였다. 이 경우에는 변경 간격 T도 경우마다 다른 범위를 가진다. 평균은 기존 평균값에 기울기 값만큼을 더해가는 방식($m_t = m_{t-1} + slope$)으로 변경되며, 표준편차는 변화하는 평균에 따라서 일정 비율의 CV(coefficients of variation)를 곱해서 구한다. 실험에서는 각 유형별로 4가지의 CV, 즉 CV=0.05, 0.1, 0.15, 0.2에 대한 결과를 비교하였다.

- MV-type 0: $slope$=0 (안정적 고객 수요)
- MV-type 1: $T = U(15,30)$, $slope = U(-1,1)$,
- MV-type 2: $T = U(5,10)$, $slope = U(-5,5)$.

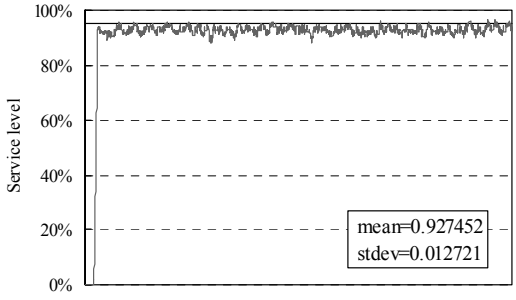
실험에 사용된 리드타임의 경우 다음의 두 가지 형태로 구분하였다.

- LT1: $L_i = 1,\ i = 1,2,3,4$,
- LT2: $L_i = 2,\ i = 1,2,3,4$.

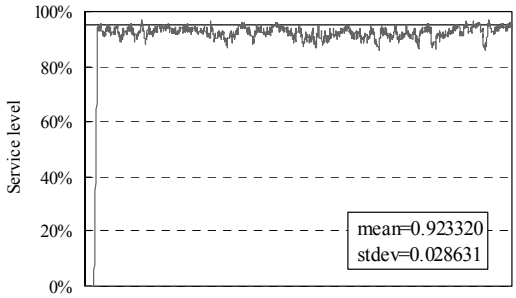
본 논문에서는 목표 서비스 수준(target service level)을 95%로 설정하였다. <Figure 2>는 LT2인 경우 본 연구에서 제시한 3가지의 MV 형태와 CV 값의 변화에 따른 서비스 수준을 나타낸 그래프이다. <Figure 2>의 결과를 분석해 보면, 본 논문에서 제안하는 재고정책이 전체적으로 주어진 목표 서비스 수준을 효과적으로 준수하고 있음을 알 수 있다. 또한 예상한대로 고객 수요가 안정적인 MV-type 0에서 서비스 수준의 평균이 가장 높고 계획기간 동안의 서비스 수준의 표준편차가 가장 작았다. 반면에 고객 수요의 불확실 정도(non-stationarity)가 가장 높은 MV-type 2에 대한 서비스 수준이 가장 떨어지며, 서비스 수준의 표준 편차 또한 가장 큰 것으로 판명되었다. 이 밖에도 고객 수요의 표준편차를 결정하는 CV(coefficients of variation)가 큰 값을 가질수록 고객 수요는 보다 불안정한 형태를 나타내게 되므로, 이에 따라 CV=0.2일 경우 실제 서비스 수준이 가장 낮았고 서비스 수준의 표준편차 또한 증가한다는 사실을 알 수 있다. 특히 가장 불확실성이 높은 MV-type 2인 경우에 대해서 본 논문에서 제안하는 재고정책은 약 2.8% 정도의 오차 내에서 서비스 수준을 충족하였음을 알 수 있으며, 이를 통해 제안된 방법론의 우수성을 입증할 수 있다.



(a) MV-type 0 with CV=0.05



(b) MV-type 1 with CV=0.1



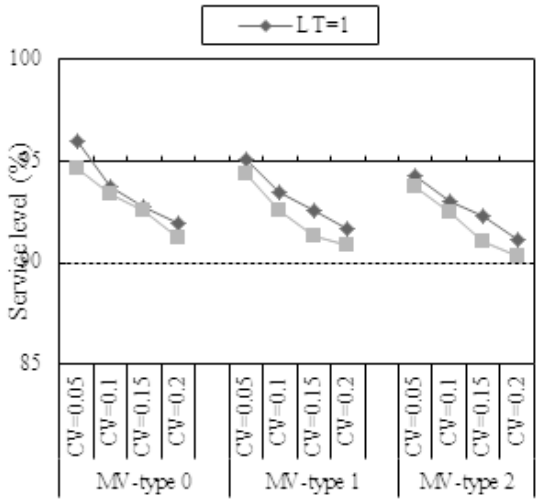
(c) MV-type 2 with CV=0.2

[Figure 2] Sample paths of service level over time

[Figure 3]는 본 논문에서 제안한 모델의 보다 구체적인 성능평가를 위해 모든 실험조건에 대한 결과를 요약해서 보여주고 있다. 그림에서 보는 바와 같이, 모든 실험조건에 대해 대체적으로 목표 서비스 수준에 근사한 결과를 보임을 알 수 있다. 보다 구체적으로 실험결과를 분석해 보면, 우선 [Figure 2]에서 살펴본 결과와 유사하게, MV-type 1의 평균 서비스 수준은 MV-type 0보다, 그리고 MV-type 2의 평균 서비스 수준은 MV-type 1보다 목표 서비스 수준과의 차이가 크다는 사실을 알 수 있다. 이러한 결과는 다른 두 타

ISSN 1229-6783(Print)
ISSN 2288-1484(Online)

입에 비하여 짧은 수요 분포 변경 간격과 큰 변동폭을 갖는 평균 수요로 인하여 MV-type 2하에서의 고객 수요의 불확실 정도(non-stationarity)가 상대적으로 크기 때문에 주어진 목표 서비스 수준을 충족시키기에 많은 어려움이 발생하였을 것으로 예상된다. 또한 이러한 현상은 리드타임의 변화에 따르는 실험에서도 유사하게 나타남을 알 수 있다. 즉 리드타임이 길수록 목표 서비스 수준을 만족시키는 정도가 낮아지며, 이는 리드타임이 길어질수록 고객 수요에 대한 불확실성이 증가하기 때문으로 해석될 수 있다. 마지막으로 CV의 변화에 따른 평균 서비스 수준의 변화를 살펴보면 주어진 수요분포 및 리드타임에 대해 CV 값이 커질수록 평균 서비스 수준이 낮아지는 관계가 성립함을 발견할 수 있다. 이는 앞에서 언급한 바와 같이 CV 값이 커질수록 마찬가지로 고객 수요의 불확실 정도가 증가하기 때문으로 해석될 수 있을 것이다.



[Figure 3] Summary of simulation results

4. 결론

기존의 공급망 문제에 대한 해법들은 현실상황에 부합하지 못하거나 변화하는 기업환경에 적절히 대응하지 못하여 실제 기업의 경쟁력 향상에 실질적인 도움을 제공하기에는 미흡한 측면이 많았다. 본 연구는 기존의 공급망 관련 연구에서 해결하지 못하고 있는 불안정한 고객수요하의 다단계 공급망에서의 서비스 수준 충족을 위해 강화학습 기법을 이용하는 새로운 적응형 재고통제 정책을 제안하는 것이며, 연구 결과를 효과적으로 활용한다면 보다 현실적으로 사용가능한 재고통제 정책을 개발할 수 있는 토대를 마련할 수 있을 것으로 기대된다. 추후 연구로는 보다 현실적인 공급망 관리를 위해 재고비용을 함께 고려하는 것이다. 즉 주어진 서비스 수준을 제약식으로 하고, 이러한 제약조건 하에서 목적함수로 전체 공급망에서의 재고유지비용의 총합을 최소화하는 것을 목적으로 하는 공급망 분배계획 방안에 대해 연구하고자 한다.

5. References

[1] Axsater, S., Inventory control, Springer (2006).
[2] Diks, E.B., Kok, A.G. de and Lagodimos, A.G., "Multi-echelon systems: a service measure perspective", European Journal of Operational Research, Vol.95 (1996) : pp. 241-263.
[3] Forteus, E., Foundations of stochastic inventory theory, Stanford University Press (2002).
[4] Gavirneni, S. and Tayur, S., "An efficient procedure for non-stationary inventory control", IIE Transactions, Vol.33 (2001) : pp. 83-89.
[5] Jiang, C. and Sheng, Z., "Case-based reinforcement learning for dynamic inventory control in a multi-agent supply-chain system", Expert Systems with Applications, Vol.36 (2009) : pp. 6520-6526.
[6] Kim, C.O., Kwon, I.H. and Baek, J.G., "Asynchronous action-reward learning for nonstationary serial supply chain inventory control", Applied Intelligence, Vol.28, No.1 (2008) : pp. 1-16.
[7] Kwon, I.H., "Distribution planning for a supply chain to satisfy target service level with non-stationary demand", Journal of the Korean Society of Supply Chain Management, Vol.10, No.2 (2010) : pp. 81-90.
[8] Kwon, I.H., Kim, C.O., Jun, J. and Lee, J.H., "Case-based myopic reinforcement learning for satisfying target service level in supply chain", Expert Systems with Applications, Vol.35 (2008) : pp. 389-397.
[9] Simchi-Levi, D., Kaminsky, P. and Simchi-Levi, E., Designing and Managing the

Supply Chain, McGraw-Hill (2008).
[10] Sutton, R.S. and Barto, A.G., Reinforcement learning, MIT Press (1998).
[11] Tersine, R.J., Principles of inventory and materials management, Prentice-Hall (1994).
[12] Vollmann, T.E., Berry, W.L., Whybark, D.C. and Jacobs, F.R., Manufacturing planning and control for supply chain management, McGraw-Hill (2011).
[13] Zipkin, P.H., Foundation of inventory management, McGraw-Hill (2000).

저 자 소 개

권 익 현

고려대학교 산업공학과에서 학사, 석사 및 박사학위를 취득하였다. 미국 University of Illinois at Urbana-Champaign에서 박사후 연구원으로 근무한 바 있다. 현재 인제대학교 산업경영공학과 조교수로 재직 중에 있다. 주요 관심분야는 물류 및 공급망관리, 생산계획 및 통제, 서비스 사이언스 등이다.

주소: 경남 김해시 인제로 197 인제대학교
산업경영공학과