Coordination of Emergent Demand Changes via Value-Based Negotiation for Supply Chain Management (Student Abstract)

Takumu Shimizu[1,2], **Ryota Higa**[2,3], **Katsuhide Fujita**[1,2], **Shinji Nakadai**[3,4]
[1] Tokyo University of Agriculture and Technology, Japan
[2] National Institute of Advanced Industrial Science and Technology(AIST), Japan
[3] NEC Corporation, Japan
[4] Intent Exchange, Inc.
shimizu@katfuji.lab.tuat.ac.jp, r-higaryouta@nec.com, katfuji@cc.tuat.ac.jp, nakadai@intent-exchange.com

Abstract

We propose an automated negotiation for a reinforcement learning agent to adapt the agent to unexpected situations such as demand changes in supply chain management (SCM). Existing studies that consider reinforcement learning and SCM assume a centralized environment where the coordination of chain components is hierarchical rather than through negotiations between agents. This study focused on a negotiation agent that considered the value function of reinforcement learning for SCM as its utility function in automated negotiation. We demonstrated that the proposed approach could avoid inventory shortages under increased demand requests from the terminal customer.

Introduction

In supply chain management (SCM), fulfilling promises made to suppliers and customers, such as meeting their demands and avoiding late deliveries, is important. However, during the trade of products in companies, it is common for planned order quantities to change later because of emergent sudden changes in SCM (e.g., boominess and depression). Thus, negotiations are conducted among humans to adjust plans in the real world to deal with such changes. However, existing studies related to SCM and reinforcement learning (RL) assume a centralized environment where the coordination of chain components is hierarchical rather than negotiations between agents(Gijsbrechts et al. 2022). In this study, we propose an automated negotiation approach that can help the RL agent adapt to situations with emergent changes and coordinate in SCM. The experimental results demonstrate that the proposed approach can avoid inventory shortages in situations involving demand change requests.

SCM Framework via Negotiation

Figure 1 shows an automated negotiation-based framework for a four-layer serial supply chain network assumed in this study. The terminal supplier (TS) is an agent with an infinite supply source, and the terminal customer (TC) makes orders based on demand determined from a certain distribution. The supplier (S) and customer (C) aim to minimize the inventory quantity I_t^i and to maximize the shipping quantity

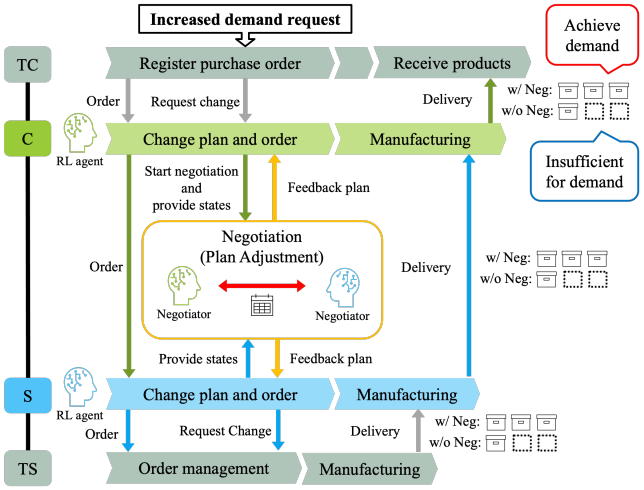

Figure 1: Automated negotiation-based framework for a four-layer serial supply chain network. The agents can start negotiations when needed. Negotiation provides feedback on plans to deal with sudden changes in demand.

D_t^i, where $t \in \{0, 1, ..., T\}$ is the current simulation step and T is the number of steps in one episode, and $i \in \{\mathrm{S}, \mathrm{C}\}$ is the agent. The agents have the arrival plan $\mathcal{O}_t^i$ and the shipping plan $\mathcal{D}_t^i$, which manage the future quantity of arrivals and shipments until after T_F steps.

Negotiations about the shipping plan between S and C are conducted to coordinate the changes, as shown in Figure 1. The agent aim to maximize the following achievement rate (AR) of the demand of the downstream agent:

$$\mathrm{AR} = N^{\mathrm{achieve}}/T, \tag{1}$$

N^{achieve} is the number of steps with sufficient inventory.

A negotiation procedure is developed based on the alternating offers protocol (AOP). In addition to the actions defined in the AOP, the agent selects the action of initiating negotiation up to once per step. This procedure can be considered as multitime repeated negotiations. The negotiators propose plans to each other until an agreement or a deadline is reached. Because the plan consists of the quantities for each step, the set of all possible bids is given by $\Omega = \prod_{j=t}^{T_F} \mathcal{Q}$, where $\mathcal{Q} = \{0, 1, ..., |D|\}$. $|D|$ is the maxi-

mum quantity shipped per step. The agreed plan is fed back to the RL agent, which updates its internal plan accordingly.

Reinforcement Learning Agent Coordination via Value-Based Negotiation

Policy based Reinforcement Learning for SCN The supplier and customer learn planning policy using RL. The agents update the internal plans based on the state $s_t^i := (\omega, \tilde{s}_t^i)$, where ω is the shared plan and $\tilde{s}_t^i$ consists of internal plans and I_t^i. Because of the use of proximal policy optimization to learn parameters in this study, the policy function of the agent i is denoted by $\pi_\theta^i(a_t^i|s_t^i)$ and the value function is denoted by $V_\theta^i(s_t^i)$. Because learning occurs without sharing rewards or parameters between agents during training, both agents' policies are present in the environment, and their models are learned alternately.

Value-Based Utility Function The utility function is defined as $U^i(\omega) := \hat{V}^i(\omega; \tilde{s}_t^i)$, where $\hat{V}^i(\omega; \tilde{s}_t^i)$ is the expected reward if the agents agreed upon ω. $\hat{V}^i(\omega; \tilde{s}_t^i)$ is calculated from multiple internal simulations, performed in the environment with the states of the agents in step t and the same TC's demand as the training environment. After N^{sim} internal simulations, the expected reward is calculated as $\hat{V}^i(s_t^i) = \frac{1}{N^{\mathrm{sim}}} \sum_{n=1}^{N^{\mathrm{sim}}} \left[\sum_{j=0}^{T-t} \gamma^k r_{t+j} | s_t^i \right]$.

Negotiation Strategies The offering and acceptance strategies are determined by a classical linear concession function with a value-based utility function[1]. The strategy that determines the start of the negotiations is defined as $\Delta^i(s_t^i) := \frac{1}{T}(V_\theta^i(s_t^i) - \hat{V}^i(s_t^i))$; when $\Delta^i(s_t^i) > \tau$, the agent determines that a negotiation is required. Because considering all possible bids in a negotiation is difficult, we considered a bid extraction strategy from Ω. The agent generates neighborhood bids by varying some dates and quantities in its optimal plan: $\mathcal{D}_t^{\mathrm{S}}$ and $\mathcal{O}_t^{\mathrm{C}}$.

Experimental Results and Discussion

Experimental Settings The number of steps (T) in the simulation was 30. The negotiation deadline was 50 rounds. The experiments considered four scenarios, depending on the variation in demand for the terminal customer. Based on the naive scenario consisting of steady demand, we prepared three other scenarios with dynamic changes. *Decrease(small)* and *Increase(small)* were scenarios where the demand decreased or increased in the short term. *Increase (large)* was a scenario where the demand increased periodically and the changes were relatively large.

We compared the proposed method with the following agents: **PrePV-Nego**: Pretrained policy and value-based negotiation. This policy was trained in a naive scenario. **Zero-Shot**: Pretrained policy trained in the naive scenario and used without retraining. **Base-Stock**: Base stock policy using base-stock levels ideal in the naive scenario. **Ideal**: Base stock policy using a base stock level that minimizes the inventory quantity without causing inventory shortages for customers and suppliers.

[1]The agent selects the offer that has a utility value near the target utility value $\hat{U}^i(\omega_k^i)$ calculated as $\hat{U}^i(\omega_k^i) = \hat{U}_{\max} - (\hat{U}_{\max} - \hat{U}_{\min})(\frac{k}{K})$ in the negotiation round $k \in \{0, 1, ..., K\}$.

	Decrease (small)		Increase (small)		Increase (large)	
	S	C	S	C	S	C
PrePV-Nego	**1.00**	**1.00**	**1.00**	**1.00**	**1.00**	**1.00**
Zero-Shot	0.73	0.97	0.80	0.87	0.63	0.43
Base-Stock	**1.00**	**1.00**	0.90	0.83	0.80	0.67
Ideal	**1.00**	**1.00**	**1.00**	**1.00**	**1.00**	**1.00**

Table 1: Achievement rate of the supplier (S) and customer (C) for each scenario. The achievement rate was calculated using the equation (1).

Experimental Results The pretrained policy used in PrePV-Nego and Zero-Shot achieved the demand for TC and C for all 30 steps in the naive scenario. Table 1 shows the achievement rates of each method in each scenario in the adaptation to different scenarios between the training and test phases. Whereas Zero-Shot and Base-Stock resulted in achievement rates of less than 1.0, PrePV-Nego achieved the highest achievement rates in all scenarios.

Discussion PrePV-Nego fulfilled the demand in all scenarios. The first reason was that the proposed method made appropriate plans for the inventory management information through negotiations. PrePV-Nego's shipping quantity was on average 33% and the inventory quantity was on average 52% higher than those of Zero-Shot, indicating that negotiations increased the quantity of product handled to avoid inventory shortages. The second reason was that the negotiations were conducted at the appropriate time. The negotiation start strategy enabled the agents to detect changes in TC demand and deal with them via negotiations at an early stage. In addition, if the result of the negotiation was incomplete, the agents repeated negotiations until a better plan was reached, which enabled dealing with large changes. In the *Increase(large)*, there was a tendency to gradually modify the plan from $t = 1$ at the beginning of the simulation to three consecutive negotiation steps.

Conclusion

This study proposes an automated value-based negotiation that can coordinate the agent in unexpected situations in SCM. Experimental results revealed that the proposed approach successfully received orders from the downstream agents under increased demand requests in the SCM.

References

Gijsbrechts, J.; Boute, R. N.; Van Mieghem, J. A.; and Zhang, D. J. 2022. Can deep reinforcement learning improve inventory management? Performance on lost sales, dual-sourcing, and multi-echelon problems. *Manufacturing & Service Operations Management*, 24(3): 1349–1368.