Proceedings of the 2024 Winter Simulation Conference
H. Lam, E. Azar, D. Batur, S. Gao, W. Xie, S. R. Hunter, and M. D. Rossetti, eds.

AN OPTIMAL CONSOLIDATION POLICY FOR VEHICLE ROUTING PLANS IN LAST-MILE DELIVERY: A DEEP Q-LEARNING APPROACH

Seokgi Lee[1], Hyeong Suk Na[2], and Yooneun Lee[3]

[1]Rayen School of Eng., Youngstown State University, Youngstown, OH, USA
[2]Dept. of Industrial and Systems Eng., University of Missouri, Columbia, MO, USA
[3]Dept. of Engineering Mgmt., Systems and Technology, University of Dayton, Dayton, OH, USA

ABSTRACT

As a result of dynamic changes in consumer shopping patterns such as purchasing more items more frequently from online retailers, large-scale online markets have a high split rate of orders. Consequently, consolidating and delivering split orders from the same customer has been addressed as a major problem in leveraging green logistics in online retailing. In this research, the last-mile logistics situations are analyzed to address two interrelated questions: (i) what would be the optimal order and shipment consolidation policy considering expected operational and economic efficiency in last-mile delivery, and (ii) what would be the resulting vehicle routing plan that improves operational and service performance. We develop an integrated decision-making framework that combines Deep Q-Network reinforcement learning with soft-update and feedback control to simultaneously optimize vehicle routing plans and consolidation policies. Experimental results show that the optimal consolidation policy improves transportation efficiency compared to other forced or non-forced transportation strategies.

1 INTRODUCTION

In this research, we focus on the order fulfillment and shipment consolidation policy and its associated vehicle routing decisions with the time window for an e-retailer business. Unlike previous consolidation studies, we propose a novel dynamic control algorithm combined with deep Q-learning (DCDQN) approach to leverage both order fulfillment and shipping consolidation in last-mile delivery under an e-commerce environment. While the existing literature addresses the trade-off between the fixed cost and the holding cost, in our setting, orders are held for at most a few hours only, which means that the extra holding cost due to waiting may be negligible, and the main trade-off lies in between the fixed cost and the additional speed-up cost due to expedited shipment. Although expedited shipment has been discussed in the literature on inventory management and supply chain risk management, it has not been addressed in the context of order fulfillment and shipping consolidation using the DQN approach. To address these gaps in our knowledge regarding order fulfillment and shipment consolidation in last-mile delivery services, our study addresses the following questions:

- How are multiple items in each order from the same customer collected from multiple warehouses?
- How are multiple shipments of the same customer merged to maximize the customer rewards?
- When should the consolidated orders be dispatched to minimize the total delivery cost?

2 SYSTEM ARCHITECTURE

To perform order and shipment consolidation, we propose a system consisting of three interactive modules designed to handle delivery orders across a predefined finite planning horizon. As a gateway to the decision-making framework, the Order Execution System (OES) manages the real-time arrival of delivery requests and forwards them to the Learning Simulator (LS). The LS, in turn, constructs vehicle routing plans and makes decisions regarding shipment and order consolidation over the planning horizon. Various operational scenarios for vehicle routing plans with shipment and order consolidation, along with their expected system rewards, are then relayed to the Learning Engine (LE), which updates optimal actions for order and shipment consolidation based on the current system state. For effective learning, the system state is designed as five elements: (i) total number of delivery requests planned over the planning horizon, (ii) total expected travel distance of all vehicles, (iii) sum of earliness or tardiness of delivery orders, which estimates the overall congestion level of current delivery plans, (iv) average inter-arrival time of delivery requests, which characterizes the arrival process of delivery orders, and (v) truck utilization. Technically, the LE runs the learning algorithm using Deep Q-learning with soft update, and the selected action by the LE is subsequently fed back to the LS, which simulates the action selected, and resultant last-mile delivery plans are implemented for an actual run.

3 NUMERICAL EXPERIMENT

The algorithmic performance of the proposed DCDQN is tested for experimental settings of different fulfillment-center locations and demand variation, which is characterized by three parameters: delivery order arrival rate, generation rates of orders which are eligible for order consolidation, and shipment consolidation. The average reward of DCDQN is compared to other consolidation strategies. For delivery orders that are eligible for ordered/shipment consolidation, (i) the Force-to-Consolidation (FTC) strategy forces them to be consolidated, and (ii) the Force-Not-to-Consolidation (FNTC) strategy forces them not to be consolidated. The average reward and standard deviation for the last 10 days are extracted and compared to other policies. This comparison is fairer as rewards during this period remain mostly steady.

To further evaluate the performance of DCDQN, we consider various experimental scenarios where diverse customer distributions are applied, in addition to the designated geographical location of the fulfillment and consolidation centers. To show robustness of the proposed DCDQN, test sets are created considering three different geographical distributions of customer locations (central, outer, and diagonal distributions) to compare consolidation strategies. As a result, the average rewards of DCDQN are relatively higher for outer and diagonal customer distributions compared to the central distribution. This is likely because the customers' locations are relatively closer to both the fulfillment and consolidation centers. Similar patterns are observed in the results of FTC and FNTC. However, it is worth noting that DCDQN performs significantly better with central distribution in terms of average rewards compared to other customer distributions. Furthermore, we demonstrate the effectiveness of the proposed algorithm using data generated from virtual deliveries made to suburban Amazon fulfillment centers in the Pittsburgh, Pennsylvania area. Overall, DCDQN yields higher average rewards, except for few cases of delivery scenarios and distributions of customer locations.

4 CONCLUSION AND FUTURE RESEARCH

The learning performance of the proposed approach is confirmed through numerical experiments using various transportation environments and delivery demand scenarios. Deep Q-Learning with soft-update using the multiple-layer neural network structure is very sensitive to the geographical location of the consolidation centers and general depots, as well as the distribution of customer locations, and therefore sometimes presents routing plans with poor learning results. When learning is performed properly, the optimal consolidation policy results in an improved reward compared to other forced (or non-forced) policies.

The proposed learning procedures need to be validated by various operational scenarios that represent real transportation and order management processes. In this paper, a limited number of instances are examined with a relatively simple state-action design to confirm model validity. Complicated transportation processes from the perspective of the logistics company and customers can be utilized to design a more accurate system state. Large-scale instances with a high dimension of a state-action structure will then be examined to address more realistic vehicle routing and order consolidation processes in last-mile delivery.