京都大学	博士（　　工　学）	氏名	NAILAH FIRDAUSIYAH
論文題目	MULTI-AGENT SIMULATION USING ADAPTIVE DYNAMIC PROGRAMMING BASED REINFORCEMENT LEARNING FOR EVALUATING JOINT DELIVERY SYSTEMS （共同配送システムを評価するためのアダプティブダイナミックプログラミングに基づく強化学習を用いたマルチエージェントシミュレーション）		

（論文内容の要旨）

This research modelled the behavior of freight carriers and Urban Consolidation Center (UCC) operator using Multi-Agent Simulation-Adaptive Dynamic Programming based Reinforcement Learning (MAS-ADP based RL) to evaluate the city logistics measure of Joint Delivery Systems (JDS) under an uncertain environment. The results indicated that the MAS-ADP based RL is better to adapt with the changing environment in city logistics and performs better than the MAS-Q-learning thus increasing the accuracy of decision making of freight carriers and UCC operator. Under the uncertain environment of city logistics, the MAS-ADP based RL could suggest the best decision of delivery method for freight carriers, the profitable UCC fee for the UCC operator and eventually reduce the negative environmental impacts.

Chapter 1 Introduction

Chapter 1 describes the general introduction of city logistics systems including definition, visions, and stakeholders involved in the systems. JDS with UCC, truck parking policy, combined policies implementation, and smart city logistics are introduced as examples of sustainable city logistics policies. The research motivation for developing learning models within MAS on city logistics problem is explained. It leads to the objectives of this research. Finally, some research contributions and benefits are listed in this chapter.

Chapter 2 Literature Review

Chapter 2 reviews the literatures related to the study. Markov decision processes as the defined approach of MAS are reviewed in this chapter, followed by its solution methods such as dynamic programming and RL. The current literatures about ADP based RL and Q-learning are examined in more details as they are used as the primary models in this research. Applications using ADP and Q-learning in MAS in the past literatures are also discussed in this chapter.

Chapter 3 Learning Model Development

Chapter 3 mainly describes the model development process. Starting with the description of agents considered in this research, agent's behavior and learning process using MAS-ADP based RL and MAS-Q-learning has been described. The use of Vehicle Routing Problem with Soft-Time Windows (VRPSTW) to optimize the agent's delivery routing problems is also described. Chapter 3 gives models used to calculate environmental emissions.

Chapter 4 Learning model evaluation in the hypothetical network

Chapter 4 applies MAS-ADP based RL and MAS-Q-learning models in the hypothetical network, which is also used to study the performance of the two learning models under different environment settings of simulations. The sustainability of JDS policy measures with UCC is also investigated in chapter 4 followed by a discussion on the benefits of learning for freight carriers and UCC operator under some given future scenarios of JDS with UCC. This chapter supports the research benefits and contributions described in chapter 1 with simulation results.

Chapter 5 Learning model application for evaluating JDS in relation with the different locations of UCC in Yokohama network

The results of chapter 4 are used as benchmark for chapter 5 for evaluating the performance of the developed MAS-ADP based RL and MAS-Q-learning models in a real network of Yokohama. The purpose of chapter 5 is to evaluate the policy measures of JDS in relation with the different

locations of UCCs on Yokohama network, Japan using MAS-ADP based RL.

Chapter 6 Conclusion

Chapter 6 summarizes the conclusions of this research as listed;

1. This research developed the learning models for freight carriers and UCC operator using MAS-ADP based RL as an on-policy learning method to evaluate the policy of JDS under an uncertain environment of city logistics.
2. The simulation results in hypothetical network and Yokohama network indicated that the MAS-ADP based RL is better to adapt with the changing environment in city logistics and performs better than the off-policy based algorithm such as MAS-Q-learning thus increasing the accuracy of decision making of freight carriers and UCC operator, and reducing the negative impact of the environment.
3. This research also shows that the “learning” is essential in the decision-making process of the agent, especially when the agent needs to interact within the multi-agent systems with fluctuating and changing environment such as city logistics. By learning the value received from the environment, learning agents could take more accurate decisions thus optimize their objective. Under the uncertain environment of city logistics, the MAS-ADP based RL increased the benefits of JDS with UCC by suggesting the best delivery method for freight carriers, the profitable UCC fee for the UCC operator and eventually reduce the negative environmental impacts.