A Study on the Application of
Multi Agent Systems for Managing Returns

Dual Degree Dissertation

Stage I Report

Submitted in partial fulfillment of the requirements for the degree of

Bachelor of Technology and Master of Technology by

Devang Thakkar (13D100007)

Under the supervision of

Prof. A. Subash Babu

DEPARTMENT OF MECHANICAL ENGINEERING

INDIAN INSTITUTE OF TECHNOLOGY, BOMBAY

Date of Submission: 12/10/2017

Contents

List of Figures

List of Tables

Abstract

Managing product returns is an increasingly important task for online marketplaces, primarily due to the volume of goods involved. Besides aiming to reduce returns, a remedial measure in this case is to optimize the reverse logistics process so as to reduce the resulting losses. Distribution centres associated with these firms are usually located outside city limits due to size and cost restrictions. A novel method is proposed where a mobile 'mother-ship' acts as a temporary accumulator so that the operational and opportunity cost of individual collectors, the 'daughter-ships', going back to the distribution centre is reduced without having to invest in storage facilities within the city. Future work will involve implementing a multi agent system for this model and organizing the schedule using multi-agent reinforcement learning algorithms.

Chapter 1

Introduction

1.1 Background

In today's world, online marketplaces are becoming more and more ubiquitous, shifting a great deal of consumer transactions from the traditional brick and mortar stores to these seemingly infinite supermarkets. More people are getting connected to the internet each day and the increasing ease of conducting a transaction on these marketplaces - particularly, the reassuring touch lent by post-delivery payment options - has lead to a gargantuan rise in orders placed on these marketplaces as compared to the previous decade. However, these marketplaces do not offer the in-person shopping experience that a good percentage of the population is used to. Thus, there has to be a more potent reason for the flux of customers away from the outlets to their device screens.

One of the most important trends responsible for the growth of online shopping is the shifting of the economy from goods to services - online retailing or 'e-tailing' marketplaces are no longer responsible for only delivering the product that the consumer needs. The entire process - right from the moment the user lands onto the web page to the comparison, selection and ordering of the product till the moment the product is delivered satisfactorily - is the product now. This has brought about a revolution in the online marketplace sector. Not only are firms improving their avenues of user interaction - primarily websites and applications, but they are also working hard to optimize their delivery algorithms, down to the last mile, while allowing for extremely convenient and rapid returns.

1.2 The Need for Returns Management

In order to attract more customers, e-tailers are concentrating mainly on improving their forward supply chain logistics. Since the advent of online marketplaces, reverse logistics has always played the second fiddle since it was not considered important or common enough. Today, it's an important area because effective returns management can lead to a definite positive impact on the firm's finances, primarily due to the volume involved [Guide Jr et al., 2006] and boost customer relations, leading to increased transactions and more revenue. Traditionally, returns management has focused only to increasing the efficiency of the chain of events. This has now given way to a more complex process with activities such as gatekeeping that remove returned products with negative expected revenue from the process at the earliest in order to minimize losses.

1.3 The Problem

For decades, the research on Supply Chain Management has been focused on the issues of the forward supply chain ranging from negotiations with suppliers at various stages to the optimization of schedule for delivery. However, given the cut-throat competition and the increasing costs of losing a consumer, focus has started shifting towards the optimization of the returns management. Even so, there hasn't been a colossal change in the manner with which research on reverse supply chain is conducted; it is still considered either as an appendage of the forward supply chain or a response to environmental regulations.

This thesis attempts to break free of the traditional model of management of returns and create a new model for collection of returned products. A novel mothership system of mobile collector and accumulator agents is implemented in order to maximize coverage, improve customer service and reduce operational costs. As noted clearly in literature, the most expensive segment of a product delivery is the 'last mile' zone ([Boyer et al., 2009]) since it requires personalized scheduling for each of the customers according to their location. This 'last mile' then becomes the 'first mile' in the case of returns management, and our key topic of interest in this work.

In the past, classical literature has mainly focused on heuristic approaches ([Little et al., 1963]; [Bellman, 1962]) to determine the optimal solution to such problems, with many scholars opting for evolutionary algorithms that provide nearly optimal solutions in the recent past. Another technique used to solve such a problem is Reinforcement Learning – a method that learns without any prior experience. Reinforcement learning algorithms work with an agent that has a defined goal and can interact with its environment, obtain quantitative feedback and moves towards its target state. The agent has a set of actions, of which it chooses one at a particular time step until it achieves its goal.

We model our system as a multi agent system wherein instead of going back to a distribution centre, the last mile collector reports to a larger mobile accumulator which then reports back to the distribution centre in order to reduce collection costs. The collector and the accumulator agents learn through active reinforcement learning a policy that helps them decide what action needs to be taken in a particular case so as to minimize operational costs and serve the customer in the minimum time possible.

1.4 Scope of Research

To the best of our knowledge, at the time of starting this research, there is no literature that describes a distributed mobile mother ship system for collection of returned products. This piece of work entails the following tasks:

- The foremost part of the investigation is concerned with the identification of the problems in returns management that may be of interest to us. The following sub-tasks ensure a holistic coverage of the topic:
  - Conduct extensive literature review on the state of returns management being implemented
  - Obtain a better understanding of the different types of returns and how they are taken care of
  - Appreciate the processes that take place once a product is collected from the customer till it is accounted for in the inventory
- The second part of this work is concerned with the creation of an agent based architecture for the implementation of our model. The following steps are needed in order to effectively implement this:
  - Define the process boundaries for our model, in order to focus the efforts

  - Implement the ISM methodology along with MICMAC analysis to:
    * obtain a list of processes involved in the model;
    * analyze the interdependencies between the processes; and
    * obtain a better understanding of the model by creating a hierarchical relation between the processes
  - Create well-defined structures to represent each of the agents and their interconnections

- The third part will be concerned with the creation of an optimal returns collection policy using reinforcement learning, comprising three stages
  - The first stage would be to teach the combination of an accumulator and a collector to learn the best way to cater to a single, stochastically appearing randomly located customer
  - The second stage would be to teach the combination of a single accumulator and multiple collectors to learn the best way to cater to a single customer for each collector, followed by multiple customers for each collector
  - The third stage would be to teach the collectors to coordinate among themselves using multi-agent reinforcement learning to optimize the collection process

The first two parts have been completed in the first phase, with the third part under progress and to be included for the second phase due to its complex nature.

Chapter 2 discusses the literature surveyed during the initial phase of the project. Each of the four main themes of this work has been provided with a sub section dedicated entirely for them.

Chapter 3 illustrates the model that has been proposed in this work. This chapter lists out the various types of processes and agents involved, and their hierarchy and interdependencies using the ISM methodology and MICMAC analysis.

Chapter 4 describes the work chalked out for the future - models, with increasing levels of complexity. to be implemented using multi agent reinforcement learning.

Chapter 2

Literature Review

2.1 Introduction

The primary problem in reverse logistics has been to define the scope of this research field – several authors have tried to shape this definition over time as part of closed loop supply chains and logistics networks. Even though the environmental concerns ensured that companies showed interest in optimizing their forward and reverse supply chain logistics starting in the early 1990s, Reverse Logistics (RL) first found a definition by [Rogers and Tibben-Lembke, 2001] as

"Reverse Logistics is the process of planning, implementing, and controlling the efficient, cost effective flow of raw materials, in process inventory, finished goods and related information from the point of consumption to the point of origin for the purpose of recapturing value or proper disposal."

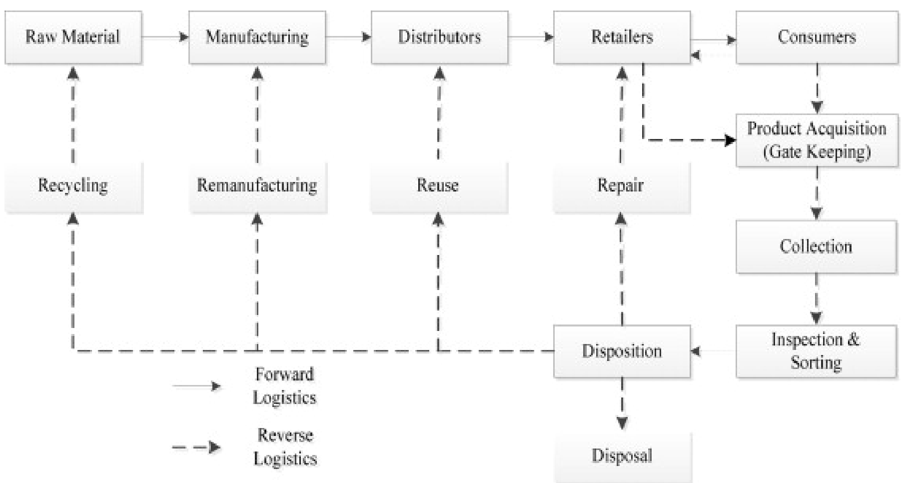

Figure 2.1: A simplistic view of forward and reverse logistics, [Agrawal et al., 2015]

Figure 2.1 describes the forward and the backward logistics process and how they are linked with each other. Reverse logistics has been a topic of increasing research due to rising interest in environmental concerns and obtaining the maximum value out of product returns ([Van Hoek, 1999]; [Wu and Dunn, 1995]; [Akdoğan and Coşkun, 2012]). Environmental regulations and reuse of disposed product parts have been the major drivers for reverse logistics, until recently [Stock and Mulki, 2009] showed that optimizing return disposition options could help organizations increase profitability.

In the earlier days, the reverse supply chain was concerned only with the 3 'R's – Recycling, Reuse and Remanufacturing. However, with the advent of distance selling and online marketplaces, Returns management has grown to be one of the major areas of RL. There are varying opinions as to which of Reverse logistics and Returns management is a subset of the other, but [Rogers and Tibben-Lembke, 2001] define Returns management as

"Returns management is the part of supply chain management that includes returns, reverse logistics, gate keeping and avoidance"

Returns have not been paid much attention to as much of the focus in the case of a majority of the industries has been concentrated on maximizing the profits from the forward supply chain ([Mollenkopf et al., 2011]). The advent of online retailing or 'e-tailing' as it is often referred to as, has led to an exponential increase in sales compared to brick and mortar outlets. This trend is hardly surprising – not only has the number of people connected to the internet has increased manifold but also the number of people opting for online purchases has seen a spike due to the ease in the payment process and the flexibility of returns in case of dissatisfaction. Besides financial concerns, enterprises are also striving to put customer relationships and loyalty at the fore of their business decisions in order to stay ahead of the competition. Customer acquisition is an expensive transaction in today's world of e-commerce and hence customer retention and e-loyalty is necessary for any business to maintain profitability. Understanding what the customers need and lowering their inhibitions provides a firm the required edge over its competition and allows it stay afloat ([Srinivasan et al., 2002]).

When there exists a possibility of returning a product in case of dissatisfaction or regret, customers are more likely to opt for the purchase in the first place. However, this ends up being a dilemma for enterprises as returns are a huge cost factor for the company, especially because of lack of optimization of the reverse product flow. [Guide Jr et al., 2006] put the commercial value of products being returned at over $100 billion per annum at an average of 6 per cent of total revenues and this cost has snowballed in the last decade with the rampant growth of the e-commerce economy. According to the definition by [Rogers and Tibben-Lembke, 2001] stated above, reverse logistics handles the process of moving products and/or packaging back from the consumption point to the site of origin. Gatekeeping refers to the process of decision making in order to restrict the number of products that are allowed to enter the reverse supply chain. Gatekeeping is an extremely critical issue since it requires the control and reduction of returns while at the same time preventing any harm to the customer satisfaction. Yet, it is the most important point in the returns process as a majority of the unnecessary cost can be cut down by screening non-utilizable products.

Due to the multi-faceted nature of this work, multiple parallel avenues were explored in order to be able to comprehend the problem better. Thus, a broad classification was deemed necessary in order to chalk out the major themes of this work. Much of the literature reviewed falls into more than one of these categories while some of it may only match the label tangentially. Table 2.1 lists out the research papers covered during the literature review stage besides categorizing them into various broad buckets concerning the theme of this research work.

Table 2.1: List of literature reviewed

Reference	Returns Management	Reverse Supply Chain	Multi Agent Systems	Reinforcement Learning
[Agrawal et al., 2015]		✓		
[Akçalı et al., 2009]		✓		
[Akchurina, 2010]			✓	✓
[Akdoğan and Coşkun, 2012]	✓	✓		
[Aras et al., 2008]		✓		
[Bakker et al., 2010]			✓	✓
[Bloembergen et al., 2015]			✓	✓
[Boyer et al., 2009]	✓	✓		
[Chaharsooghi et al., 2008]		✓		

Continuation of Table 2.1				
[Franklin and Graesser, 1996]			✓	
[Gatti, 2014]				✓
[Govindan et al., 2015]		✓		
[Griffis et al., 2012]	✓			
[Guide Jr et al., 2006]	✓			
[Harris, 2010]	✓			
[Jiang and Sheng, 2009]		✓	✓	✓
[Kovalchuk, 2009]		✓	✓	
[Min et al., 2006]		✓		
[Mollenkopf et al., 2011]	✓			
[Monostori et al., 2006]			✓	
[Panait and Luke, 2005]			✓	
[Rogers and Tibben-Lembke, 2001]	✓	✓		
[Russell et al., 1995]				✓
[Sangwan, 2017]		✓		
[Shen et al., 2006]			✓	
[Srinivasan et al., 2002]	✓			
[Stock and Mulki, 2009]	✓			
[Stock et al., 2006]	✓			
[Stockheim et al., 2003]		✓		✓
[Sutton and Barto, 1998]				✓
[Tanaka and Yoshida, 1999]				✓
[Van Hoek, 1999]		✓		
[Veluscek, 2016]		✓		
[Walsh et al., 2014]	✓			
[Wooldridge and Jennings, 1994]			✓	
[Wooldridge and Jennings, 1995]			✓	
[Wu and Dunn, 1995]		✓		
[Zolfpour-Arokhlo et al., 2014]			✓	✓

2.2 Returns Management

The first step towards solving the product returns problem is to understand the importance of optimizing product returns and the problems involved with it. In this section, we shall first understand the various ways in which returns have been classified in literature, followed by a brief overview of the returns process. In their seminal work, [Rogers and Tibben-Lembke, 2001] define five major categories of returns. The major focus of our work will be limited to 'Consumer returns', however, all the categories of returns have been explained below for the sake of completeness:

1. *Consumer returns:* Consumer returns due to problems with size or defects are reported to be the most common reasons for returns, but there are many reasons including buyer's regret which may not be recorded due to falsified returns reports. This can be seen from the survey conducted by [Harris, 2010], wherein he reports that up to 92 per cent of the consumers reported having returned a product fraudulently. Consumer returns are important because a customer friendly returns policy helps build customer loyalty which in turn leads to higher sales down the line.

2. *Marketing Returns:* Marketing returns include the backward flow of goods from a forward position (but not the final consumer) in the supply chain. This may happen due to a variety of reasons, some of which are listed below:

    - Close out – This happen when the distributor or the retailer does not wish to sell the product any longer and wants to ship it backwards up the supply chain – this may include customer returns or sample products.
    - Buy out – This is an oft-used marketing technique wherein a manufacturer buys out the competitor's products in order to obtain shelf space in highly competitive markets.
    - Job out – This is when some seasonal merchandise has failed to sell and need to be gotten rid of.

3. *Asset Returns:* Asset returns include the back propagation of assets that have been moved forward in the supply chain and have no purpose at that position in the supply chain. Most of the assets that need to be sent back are some kind of large scale packaging items such as reusable containers or large cartons.

4. *Product recalls:* Product recalls are made by firms in cases where they believe that their product might not be of optimal quality or might be a safety concern for the consumer. This might not concern most of the retail sector, but is a major concern for fields such as the automobile or food products industry that are more susceptible to such recalls.

5. *Environmental reasons:* Due to increasing pollution concerns, governments now have strict regulations in place in order to ensure sanitary disposal of toxic material wherein the waste and/or the packaging is sent back up to the manufacturer for recycling and treatment.

[Stock et al., 2006] discuss two important concepts – (i) understanding the types of returns, and (ii) enumerating the steps involved in the returns process. It classifies returns into two essential categories – controllable and uncontrollable returns. Controllable returns deal with those returns which could have been avoided by resolving the respective problem before its incidence. Controllable returns occur due to damage to products, problems with sizes and intrinsic product defects, thus the optimization of forward logistics would be of little help in avoiding those returns. Some of the ways to reduce these returns would be to:

- Improve packaging and handling to reduce in-transit damages to products
- Accurately represent sizes and colors in product descriptions to avoid misplaced expectations
- Stricter quality checks before the product leaves the manufacturing shop
- Better shop floor automation to avoid delivery of incorrect goods

On the other hand, uncontrollable returns are those that are not a result of a shortcoming on the part of the supplier and hence are inevitable. It is for these returns that optimal systems for reverse logistics need to be developed. The goal with regards to these is to place them efficiently into the backward supply chain after an optimal level of gatekeeping. This would ensure that not only a small number of products enter the reverse logistics network but also that a good fraction of it can be refurbished, repackaged and sent back into the forward supply chain. Even those goods that may not be sold directly can be remanufactured, used for warranty repair purposes or sent back to original equipment manufacturers for dismantling and reuse.

This paper also divides the product returns process into five stages:

- *Receive:* Procuring the items back from the customer is one of the biggest problems associated with the returns process. In forward logistics phase, goods are shipped out in a fairly optimised manner, down up to the last mile. However, this is not the case for returned items. Present in all shapes and sizes, returned items need to be aggregated and shipped back as quickly as possible so that it may be discarded or re-enter the forward supply chain at the earliest.

- *Sort and Stage:* The returned goods need to be sorted in order to determine the further processes that it would undergo. This process is usually merged with the receipt process, and is expected to be completed within two to three days of the product collection.

- *Process:* This stage includes the sorting of returned goods into different categories according to the original vendor or inventory re-entry. This stage helps to avoid fraudulent returns, a major loss to the firm ([Harris, 2010]).

- *Analyze:* This is where the technical know-how comes into the picture. The employees need to be able to understand the difference between allowable and non-allowable returns, the various repairing or remanufacturing opportunities possible and their respective profit potential. Once the legitimacy of the returns has been confirmed, credits may now be offered to the customer.

- *Support:* Having decided the disposition of the returned products, the items are distributed into their respective channels. Time is of utmost importance, regardless of which channel the item belongs to – (i) the faster the item re-enters the supply chain, lower the holding costs, (ii) vendor return windows can be extremely short, and (iii) the faster the process takes place, the lesser is the cash held up in the returns inventory.

[Walsh et al., 2014] analyze the different methods which may be used to reduce consumer returns. The methods described in his paper are divided into three broad instrument categories as shown in Table 2.2:

Table 2.2: Possible instruments to reduce product returns ([Walsh et al., 2014])

Monetary	**Customer based**	**Procedural**
Return penalty	Virtual try-on	Safety packaging
Money back guarantee	Lucid product reviews	Different return channels
Shipping costs	Detailed specifications	Ban return sinners

[de Koster Marisa P. de Brito and van de Vendel, 2002] try to understand the factors on which the decision of integrated versus isolated forward and backward flows is based on. The study is based on nine real world enterprises and the factors that affect the returns management process. One of the major debates in the literature on reverse logistics is concerned with deciding the better of the two choices – combined or separated flows. Storage for returns has infrastructure and skill requirements that are quite different compared to the forward inventory. The empirical evidence listed in this work shows that an integrated process, though less expensive due to the sharing of space, is often more time consuming and complex. On the other hand a suggestion is made for collection vehicles to be the same for collection and distribution of goods.

2.3 Reverse Supply Chain

Having understood the types of returns and how they are to be processed, there is now a need to establish an efficient architecture for the reverse logistics. Network design is an extremely important part of supply chain management – it includes setting up not only the responsibilities of each of the entities in the infrastructure, but also the manner of interaction among themselves. This paper focuses on the creation of a non-conventional network architecture, however, in order to do that, we need to look at the classical approach to the problem. [Akçalı et al., 2009] enumerate some of the possible areas in which work may be needed to create an efficient network design:

- The number, location and size of the warehouses [Aras et al., 2008]
- The association of recovery activities to the aforementioned warehouses
- The flow and transport of goods among these warehouses [Avci and Selim, 2016]

One way in which product returns may be optimised is by taking advantage of the economies of scale. Traditional methods involve using aggregation of returned products into larger shipments at consolidation centres before moving them upwards for further processing.

To understand the metrics for comparison of various methods, [Sangwan, 2017] lists out several key performance indicators for both location and collection policy decisions in Table 2.3:

Table 2.3: Key Performance Indicators (KPIs) for different policies

Location Allocation	**Collection policy**
Collection cost	Initial investment
Processing cost	Volume of returned goods
Customer satisfaction	Operating costs
Product reclamation	Customer satisfaction
Energy use	Environmental impact

In another real world problem, [Veluscek, 2016] describes the supply chain optimization process developed for Caterpillar, followed by the usage of an Ant Colony metaheuristic to find the most efficient solution. One of the drawbacks of metaheuristic algorithms is the high number of parameters involved. This work focuses on building new multi-objective optimization algorithms and allowing for the usage of Ant Colony System for industrial purposes by reducing the number of parameters involved and the time taken by 60 per cent.

[Min et al., 2006] discuss a multi-echelon network where the levels are the customer, an initial collection point and a centralized collection centre that decides the fate of the returned products after gatekeeping. The questions addressed in this work are equally important to our study:

- Where should the initial collection points be located so as to reduce the time taken to serve future return requests learning from current trends?
- Where should the secondary consolidation centres be located so as to be optimally distanced from both the initial collection points and remanufacturing/forward supply warehouses?
- How should the collection be scheduled in order to follow all time related security protocols while not compromising on service level?
- How often should the products be shipped from the primary collection centre to the secondary consolidation centre, so as to minimize shipping costs?

As can be discerned from the above list of problems, the network design problem is a multi-attribute problem such that it is not possible to optimize over all the factors involved; thus the various trade-offs are to be analyzed. Insight on solving this problem can be obtained by referring to a related body of work that faces similar trade-offs: Inventory control for supply chain management ([Govindan et al., 2015]; [Kovalchuk, 2009]; [Jiang and Sheng, 2009]; [Chaharsooghi et al., 2008]). Inventory management is an equally dynamic environment where different entities try to maximize their outcomes while ensuring uninterrupted production. A decent body of literature in this field utilizes multi-agent systems and reinforcement learning approaches that are relevant to our work.

2.4 Multi Agent Systems

The term 'agent' has been a source of great contention in the academia and the industry alike. Various scholars have used the term to define various systems based on their expectations of the agent ([Franklin and Graesser, 1996]). However in our context, we describe an agent as defined by [Wooldridge and Jennings, 1995]:

> *"An agent is a computer system that is situated in some environment, and that is capable of autonomous action in this environment to meet its design objectives"*

An agent, therefore, can be described as an independent entity that attempts to assess its environment and takes the action that promises to provide the maximum reward or move a step closer towards achieving the task specified. In order to do so, an agent needs to be able to perform its actions without requiring any intervention and it should be able to store and access its states and possible actions.

Due to the poorly articulated definition of the 'agent', it is no surprise that there exists no universally accepted definition of a 'multi-agent system'. Therefore, we broadly define a multi-agent system as an environment wherein two or more agents interact within themselves, with partial or complete access to information about the environment and each other's states and experiences, in order to achieve a certain goal. Most real world systems can be reduced to models wherein the different players may be abstractly represented by agents, and the processes they are involved in are - in essence - simulations of multi-agent systems. This can be especially useful in the case of supply chain management where we have various decision makers, represented as autonomous agents who through communication and collaboration with other agents, working in tandem to accomplish their goals.

[Russell et al., 1995] classify the various types of environments into the following contrasting categories:

- *Accessible vs. Inaccessible:* In an accessible environment, the agent always has complete information about the environment while in an inaccessible environment, the agent only has updated information about the area in which it has vision.

- *Deterministic vs. Stochastic:* In a deterministic environment, if an action is taken in a particular state, it would always lead to the same resultant state; the same can not be said about a stochastic environment.

- *Static vs. Dynamic:* In a static environment, the only changes that may occur would be due to the actions taken by the agent, without any external intervention, while a dynamic environment, similar to the real world, is not exempt from such changes.

- *Discrete vs. Continuous:* The states of an environment may be finite and distinct or continuous and infinite. Although most realistic systems would be classified as continuous, theoretical models are discretized for simplification of the problem.

Due to their ability to deal efficiently with low-probability high impact situations, multi-agent systems have been used in varied fields of manufacturing and operations research. For example, [Kovalchuk, 2009] considers the problem of countering uncertainty in pricing and auctions in supply chain management. It starts with a creation of an agent-based architecture for the supply chain followed by a comparison of a variety of algorithms for the prediction of demand and the behavior of competitors. It also goes on to discuss the functionalities of agent based models that participate in the Trading Agent Competition for Supply Chain Management (TAC SCM) and how they differ from each other. Finally, Genetic Programming and Neural Network based algorithms are proposed and tested against known benchmarks in the TAC SCM game.

As a review of the applications of agent based manufacturing, [Monostori et al., 2006] list out the various possible applications of multi-agent systems in manufacturing domains such as engineering design, network planning, process planning and resource allocation among others. It also makes the case for agent based modeling for simulation of complicated systems. Unlike conventional top down methods, an agent based simulation tries to assess the local interactions that might have an impact on the global behavior of the system. Due to its decentralized nature, multi-agent systems can tackle many traditional problems with more ease due to their ability to incorporate coordination and cooperation. The only obstacle that remains is to integrate these systems with existing traditional systems.

2.5 Reinforcement Learning

Reinforcement learning is a learning paradigm based on agent based architecture. Placed in an environment, an agent interacts with the environment in a closed loop fashion. The agent observes the state of the environment and tries to access as much information about the environment as it can, after which it analyses the information and decides to perform an action which it thinks would provide the greatest reward. The action results in a finite reward which acts as a feedback mechanism for the agent which it used to improve its estimation of the state of the environment – a good action gets greater rewards, thereby 'reinforcing' it. A simple model describing how reinforcement learning works is shown in Figure 2.2.

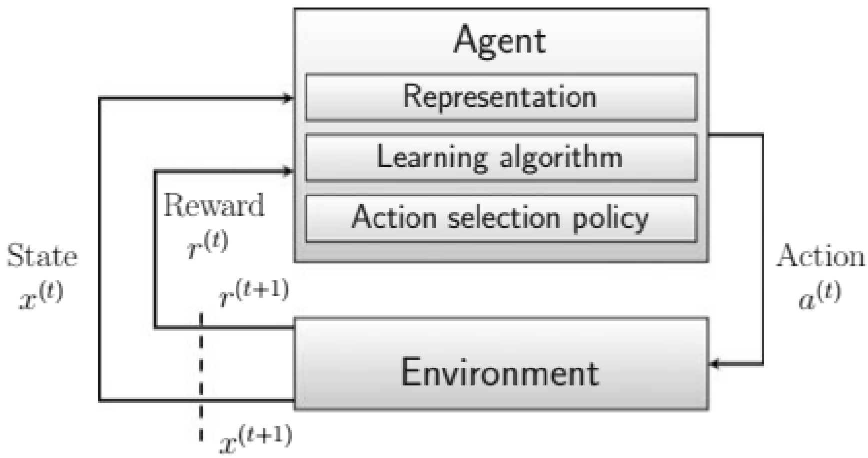

Figure 2.2: A block model of the reinforcement learning paradigm, [Gatti, 2014]

An agent does not possess any knowledge as to what a good reward is or what how far is the current action from the optimal action – it needs to try out various actions and discover the rewards offered by them ([Sutton and Barto, 1998]). As shown in Fig 1.1, the agent observes the state of the environment as x(t) at time t, takes an action a(t) based on the observation, receives a reward r(t) and the state of the environment changes to x(t+1). The agent continually updates its policy by assessing the magnitude of the reward obtained. In order to reach the optimal action, the agent follows a combination of exploratory and exploitative policies to reach not just the locally optimal action but also the globally optimal action after an adequately long learning phase. One of the ways to estimate the optimal solution is to use the Bellman equation, which is in essence, attempts to maximize the expectation of the rewards of the succeeding states. Though this application of Dynamic Programming (DP) is effective, this method requires the determination of rewards of all permissible states. A slightly less computationally intensive method is known as the Monte Carlo Method (MCM), the update rule for which is described in Equation 2.1:

$$v_{new}(\pi(s_t)) = v(s_t) + \alpha[R_t - v(s_t)] \tag{2.1}$$

where $v(s_t)$ is the value of state at time t, R_t is the reward obtained on achieving this state, and α is called the learning rate parameter. Whenever a state is encountered, its value is updated by the correction term $\alpha[R_t - v(s_t)]$ which is proportional to the difference between the received reward and the expected reward and the next state is selected by implementing the action with the maximum value. This, however, may lead only to a local optimum and not a global optimum. In order to counter this, the action with the maximum expected value is selected with a probability of (1-epsilon), with a random action chosen otherwise, a policy that is known as an epsilon greedy policy. An optimal combination of DP and MCM is called Temporal Difference Learning (TDL), a variant of which – Q Learning is the method of choice for our work, described mathematically in Equation 2.2.

$$Q(s_t, a_t) \leftarrow Q(s_t, a_t) + \alpha[r_{t+1} + \gamma \max_{a_t} Q(s_{t+1}, a_{t+1}) - Q(s_t, a_t)] \tag{2.2}$$

Reinforcement Learning has been applied to manufacturing optimization for a long time. Some avenues where this learning technique has been used are supply chain ordering ([Chaharsooghi et al., 2008]), route planning ([Zolfpour-Arokhlo et al., 2014]; [Bakker et al., 2010]), inventory control ([Jiang and Sheng, 2009]), work-floor scheduling ([Tanaka and Yoshida, 1999]) and process planning ([Mahadevan and Theocharous, 1998]). [Stockheim et al., 2003] consider the application of reinforcement learning to supply chain management wherein scheduler agents try to learn an optimal policy for accepting the requested jobs. The reinforcement learning algorithm is in charge of ranking jobs on the basis of their due date, job cost, lateness penalty and priority level, which are then executed using a scheduling agent. As expected, the algorithm outperforms conventional acceptance heuristics. [Zolfpour-Arokhlo et al., 2014] consider the problem of routing vehicles in cities of Malaysia by studying the relative importance of various parameters such as traffic, weather, road quality, time of the day and fuel capacity. The paper proposes a model using multi-agent reinforcement learning algorithms to discover the optimal routes. The model uses two groups of agents - one for estimation and one for evaluation and control. The agents may observe their local environment, work towards achieving their own local goals, and communicate and negotiate amongst themselves to obtain an understanding of the global environment and work towards the overall goal. In the end, the paper compares simulation results and experimental values and observes that the proposed approach is better than the existing approach. [Akchurina, 2010] has developed several multi agent reinforcement learning algorithms that converge to the Nash equilibrium for general cases. The algorithms developed are able to converge to a solution even when the environment is partially known, a new development in the field of games but an essential requirement for reinforcement learning.

2.6 Summary

The review of literature presented above helps us realize a few important facts. Firstly, product returns are an important part of supply chain management. The last mile problem, or in this case, the first mile problem is an important area that requires our attention due to the high costs involved. The conventional design of reverse logistics networks has greatly ignored the first mile problem by conveniently assuming drop-off points where customers are expected to drop their returned packages off. This is an example of blatant oversimplification in today's scenario - Firms are trying their best to provide faster deliveries, easier returns, and an increased level of customer satisfaction. Thus a radically new model that helps reduce operational costs while ensuring the best level of customer service is needed.

Another area of breathtakingly fast advancement in today's world is artificial intelligence. Due to their decentralized nature, multi-agent systems can take decisions by themselves, and cooperate and coordinate to work towards a common goal. Using reinforcement learning algorithms is important because it allows the system to learn with any input from the user. This is especially important in cases where there might not be many supervised inputs provided, as is the case with a lot of real-world problems. The results of the application of reinforcement learning are at par with conventional algorithms and heuristics. The real reason reinforcement learning has been chosen is because it is one of the fields that has seen an exponential rate of development in the recent past. Using Reinforcement learning in our approach allows us to keep up with the cutting edge methods in the field of learning.

Chapter 3

A Proposed Model

3.1 Introduction

Due to its complicated nature, supply chain management is an area of research where the final result, instead of being dependent on a single action, is dependent on the actions of all the components of the network. This means that we need to attempt to understand the mechanism of each of the component not just individually, but also of the network as a whole. An agent based architecture would be perfect in this case for a couple of reasons – it would provide a decentralized and scalable environment, and also allow the agents to interact and cooperate with each other. Since a majority of the functions of the components would be independent of each other, time may be saved by processing tasks in parallel, while coordinating with each other. [Shen et al., 2006] define an agent based system as:

> *"An agent based system is a loosely coupled network of problem solvers that work together to solve problems that are beyond their individual capabilities"*

It would therefore seem that a higher number of agents would lead to greater efficiency of the system. However, the advantages due to the increase in the number of agents is offset by the extra effort required for communication and the increased complexity of the architecture. Thus an optimal number of agents needs to be set up in order to maximize the efficiency.

Besides the number of agents, we also need to set up a hierarchy of agents as would be found in a real life supply chain network. We create four types of agents for a particular area being served by our system, with at least one implementation of each agent – an Accumulation agent, an Inventory agent, a Collector agent(s) and a Returns Manager agent. Thus, the architecture has now been decentralized such that each agent can take its own decisions and any number of agents may be added or removed from the system as long as the presence of one agent of each type is ensured. Each of the agents contributes towards the final goal of maximizing the profit of the enterprise in its own manner – minimize the operational cost of logistics, minimize the holding cost, minimize the customer dissatisfaction penalty. Figure 3.1 shows a UML diagram ([Bergenti and Poggi, 2000]) illustrating the relation between the agents, their individual goals and the steps taken in order to achieve the goals.

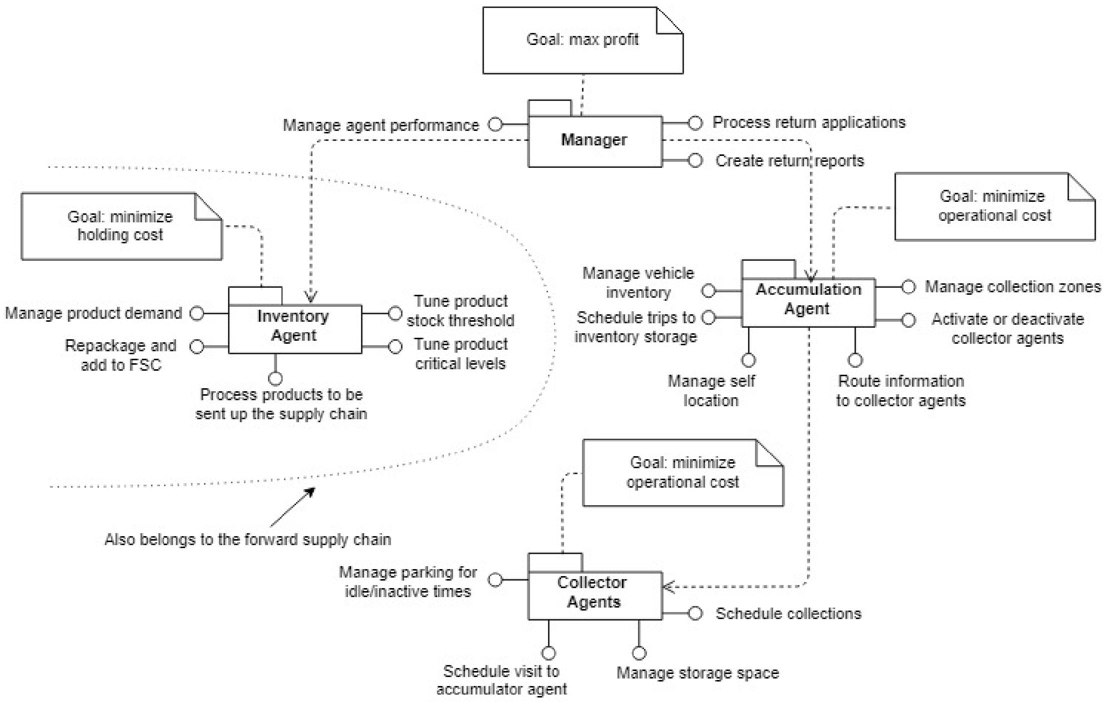

Figure 3.1: A UML model describing our agent based model

3.2 Types of Agents

3.2.1 Returns Agent

The Returns agent is the topmost authority in this hierarchy of agents and handles all interaction of the system with the environment on the customer side. Primarily, the Returns agent is responsible for the following tasks:

- Process incoming return applications and sort them in order of urgency
- Create return reports for the Accumulator agent containing spatial and temporal details for products to be collected

- Assess inventory levels for the warehouses
- Manage agent performance

The overall aim of the Returns agent is to maximize the profit after the value of the returned product has been recovered.

3.2.2 Accumulator Agent

The Accumulator agent, along with the Collector agent(s), shall form the crux of our study. We intend to try to understand the working of a mother-daughter collection mechanism through these agents. The Accumulator agent is expected to perform the tasks listed below:

- Divide up the area to be served into collection zones such that one Collector agent is responsible for one zone
- Manage self-location such that it is easily located with respect to Collector agents while moving or being placed in a permissible location
- Schedule trips to inventory location to offload collected goods
- Manage vehicle inventory – either rent vehicles as per requirement or possess a fleet of self-owned vehicles
- Route information to collector agents based on their allotted zone
- Activate or deactivate agents in times of low and high returns frequency respectively

3.2.3 Collector Agent

The Collector agents are responsible for collecting the returned products from the customer and transporting them to the Accumulator agent in our case, or the Distribution Centres(DCs) in the conventional case. Primarily, the main functions of the Collector agents are:

- Schedule return collections based on due dates provided
- Schedule return visit to Accumulator agent to unload collected products
- Manage storage space within self
- Manage self-location for idle/inactive time periods

Using such a multi-agent architecture enables us to create a versatile framework that can be reconfigured according to the requisite environment condition. A decentralized approach to the architecture allows us to distribute the tasks to agents located in different locations and permits the agents carry out independent decisions while cooperating with each other to achieve the overall goal.

3.3 Process Boundaries

The first step to solving our problem is to define the process and our areas of interest in the entire chain of events. We do this by defining process boundaries for the entire process, beginning from the placement of the order to the return of the product and its re-entry into the inventory/remanufacturing process. Of the seven processes listed below, we shall be concentrating on the last three processes only. A representative flowchart of the process boundaries has been shown in Figure 3.2

1. *Demand Trigger:* The firm lists out its products on its shopping interfaces including but not limited to desktop websites and mobile applications. The customer may place their order on the company's shopping interface and may choose to pay for it either by an online transaction or payment on delivery.

2. *Demand processing:* Once the order has been placed, the seller is informed of this. Once ready, the item may be

   - dispatched from the seller to the distribution centre; or
   - dispatched directly from the seller to the customer, depending on the firm's policies

   Most firms prefer to store an inventory of the seller's popular goods and deliver it to the customer when an order is placed.

3. *Payment process:* Firms offer their customers of paying either by online means or defer the payment post receipt of their ordered goods.

4. *Delivery process:* The delivery process under consideration starts from the distribution centre and ends at the customer. The distribution centre may send it to a city distribution centre that may directly supply it to the customer or forward it to a last-mile storage unit. This hierarchy helps to improve the flexibility of the forward supply chain and provide faster services.

5. ***Returns Collection:*** The customer files a return/exchange request. For now, we only consider returns with money back option. Once a request is filed, depending on what zone the customer belongs, a collector agent is associated with it, who procures the product from the customer. The collector agent needs to take care of the following pointers as well:

   - A collector agent may be activated or deactivated by the accumulator agent as per the long term demand pattern.
   - During its idle and deactivated times, a collector agent needs to be located in a legally approved parking zone.
   - The collector needs to report to the allocated/ nearest accumulator based on the general policies of the architecture.
   - The collector would need to report back to the accumulator either
     - when the storage space is running low; or
     - during idle times(times when collection is not possible, i.e. end of the day)

6. ***Returns Accumulation:*** For a given region, there can be a certain number of accumulators, each looking over a certain sub-region. An accumulator receives all the returned products from the collector agents. The accumulator also makes a note of the returned products (gatekeeping) and updates their status online so that the customer may be credited back the money at the earliest. The accumulator has to be located in a fixed parking slot or be moving following the traffic regulations.

7. ***Return Deposition:*** The accumulator also has to go back to the distribution centre either according to a fixed schedule or whenever the storage capacity has been utilized up to a certain extent. The accumulator may only go to the DC during inactive hours because it may be necessary for a collector agent to empty its load during the active hours.

Enumerating these processes is extremely important because it gives a direction to the study. As seen from the listing above, these processes have subprocesses that may be independent from each other. Identifying the relation between these subprocesses helps us identify the mutually exclusive subprocesses as well as those that are interdependent. The exact relation between the sub-processes can be identified using Interpretive Structural Modeling(ISM) (Watson, 1978) and MICMAC analysis, as demonstrated in the next section.

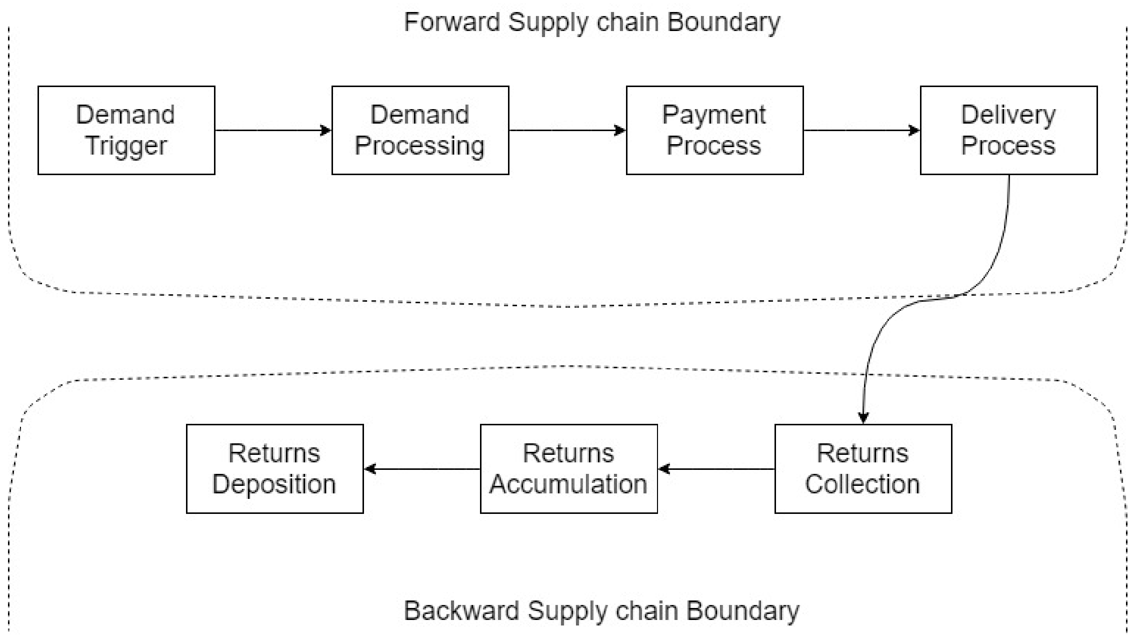

Figure 3.2: A flow chart describing the process boundaries for our problem

3.4 The Proposed System

The agents described previously form the crux of our proposed model. Having defined the process boundaries of our system, we are then faced with task of understanding the relations and hierarchy among the tasks in these processes. In order to obtain a better idea of this interdependence, we apply the ISM methodology, followed by a MICMAC analysis which is described in the next section.

We propose a mother-daughter architecture as used often in naval and aerial logistics operations.([Houghton and Beers, 1961], [Young et al., 2005]) In order to avoid the high logistics cost of a collector vehicle going back to the distribution centre at every instance of a full inventory, we propose that the collector vehicle acts as a daughter with a larger accumulator vehicle acting as the mother. An accumulator vehicle would correspond to a particular zone, subzones of which would be served by collector vehicles. We model these vehicles as agents that have information not only about their own state and about the customers in their zone, but also about their neighboring collector agents. Using this property, one possible addition to the model would be to allow a collector to cover up for a neighboring collector in times of need in order to increase the level of service. One of the important tasks is to find the optimal policy for the collector vehicle with respect to trip-planning and storage management. We are currently in the process of implementing vanilla reinforcement learning to allow the collector to learn to cater to a single customer and report to an accumulator. More complex algorithms would then be applied to cover scenarios as realistically as possible. In order to allow the collectors to cooperate among themselves and coordinate with the accumulator, we shall be using multi-agent reinforcement learning in order to find the optimal equilibrium strategy.

3.5 Application of ISM Methodology

Having a large number of contributing processes makes the analysis of a model complicated since it may not always be possible to see the nuances of the relations between tasks at the first glance. Therefore, it is necessary that a well-defined hierarchical architecture is defined for the model. Interpretive structural modeling (ISM) is a methodology that helps to understand the structure of a poorly defined mental model. ISM identifies the direct and indirect relations between processes and leads to a holistic understanding of the model.

The first step in ISM is to identify the variables involved in the system, in this case, the processes in our model. This is usually done by a literature review and a proper understanding of the system. Next, a structural self-interaction matrix (SSIM) is developed by classifying the inter-variable relations into pre-defined buckets as specified by the method. From the SSIM, an initial reachability matrix (RM) is formed by adjusting the matrix for indirect relations between processes, followed by transitivity analysis which leads to the final reachability matrix. Through this final reachability matrix, a hierarchical model is formed through partitioning of elements using antecedent and reachability sets. A schematic describing the ISM methodology has been shown in Figure 3.3

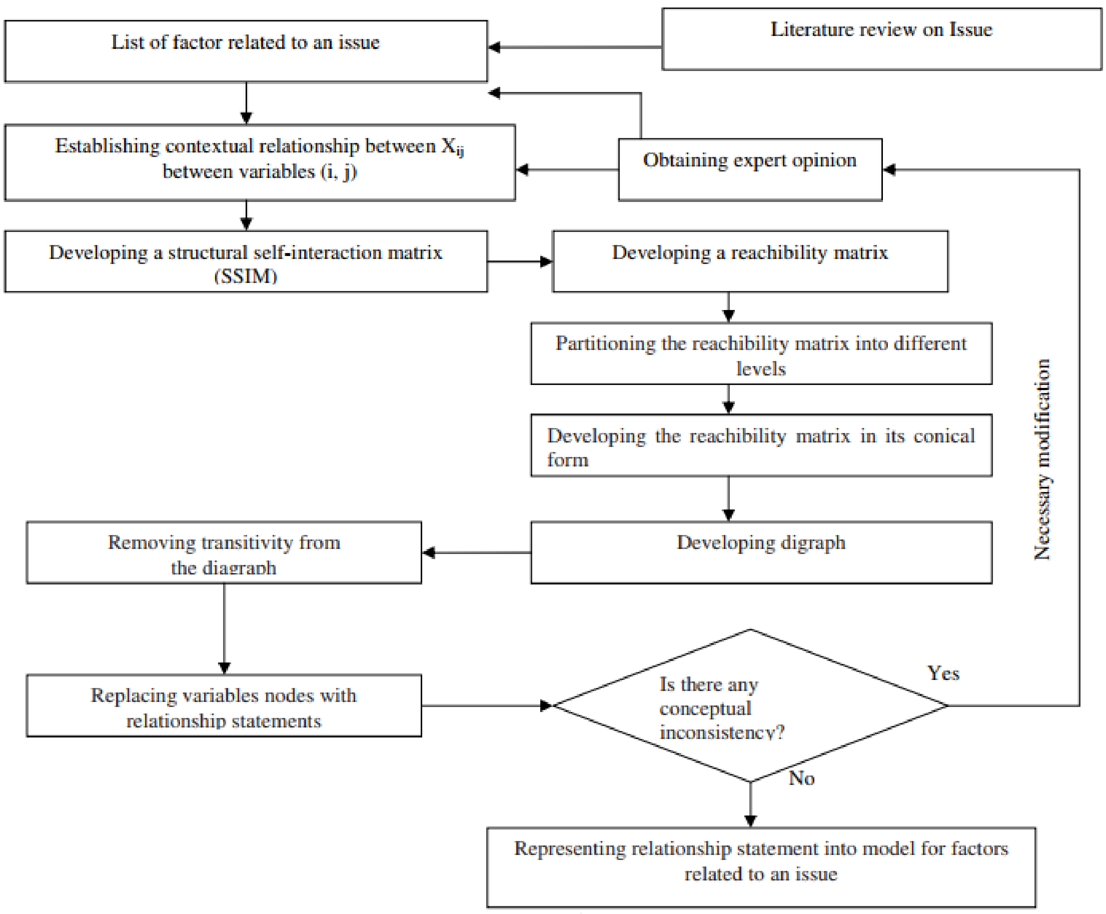

Figure 3.3: A flow chart describing the ISM methodology, [Attri et al., 2013]

3.5.1 Identification of the processes

The returns management process has been into three major sections: returns collection, returns accumulation and returns deposition, each of which have their own list of processes as listed here.

- Variables in Returns Collection:

  A1. Receive data about customer (location/due date/quantity)

  A2. Approach customer and collect returned product

  A3. Optimize collection schedule to minimize distance traveled

  A4. Manage location during idle times to optimize parking costs and collection from next customer

  A5. Manage inventory status to avoid refusal of returned product due to full inventory

  A6. Return to accumulator

- Variables in Returns Accumulation:

  B1. Perform gatekeeping for returned products

  B2. Manage location so as to optimize distance traveled by collectors, and its operational and parking costs

- Variables in Returns Deposition:

  C1. Return products to distribution centre/warehouse while ensuring enough inventory space and optimize its schedule to ensure uninterrupted service

3.5.2 Creation of a Structural Self Interaction Matrix

In order to identify the structure, we first list down the relation between the processes in an objective fashion. The SSIM is then created by noting down the relations for each pair of processes as shown in Table 3.1. For any two processes i and j, the direction of relation is given by the following symbols:

- V, if and only if the process i leads to process j
- A, if and only if the process j leads to process i
- X, if the processes i and j are codependent
- O, if the processes i and j are mutually independent

Table 3.1: SSIM matrix describing the relations between the processes involved

	A1	A2	A3	A4	A5	A6	B1	B2	C1
A1	-	V	V	O	V	O	O	V	O
A2	A	-	A	O	X	O	O	O	O
A3	A	V	-	X	A	X	O	X	O
A4	O	O	X	-	O	O	O	X	O
A5	A	X	V	O	-	X	O	O	O
A6	O	O	X	O	X	-	v	O	O
B1	O	O	O	O	O	A	-	O	O
B2	A	O	X	X	O	O	O	-	V
C1	O	O	O	O	O	O	O	A	-

3.5.3 Creation of Reachability Matrices

Once the SSIM has been constructed, an initial reachability matrix is built by finding the direct successors of processes. A cell $c_{i,j}$ is given a value 1 if its value in the SSIM is either V or X. Table 3.2 shows the initial reachability matrix obtained from the SSIM:

Table 3.2: Initial reachability matrix

	A1	A2	A3	A4	A5	A6	B1	B2	C1
A1	-	1	1	0	1	0	0	1	0
A2	0	-	0	0	1	0	0	0	0
A3	0	1	-	1	0	1	0	1	0
A4	0	0	1	-	0	0	0	1	0
A5	0	1	1	0	-	1	0	0	0
A6	0	0	1	0	1	-	1	0	0
B1	0	0	0	0	0	0	-	0	0
B2	0	0	1	1	0	0	0	-	1
C1	0	0	0	0	0	0	0	0	-

Once the initial reachability matrix has been constructed, a modified reachability matrix is built by considering the indirect relations between the processes in order to get a holistic idea of the model. Table 3.3 shows the final reachability matrix obtained by modifying the initial matrix.

Table 3.3: Final reachability matrix

	A1	A2	A3	A4	A5	A6	B1	B2	C1
A1	-	1	1	1	1	1	0	1	0
A2	0	-	1	0	1	1	0	0	0
A3	0	1	-	1	1	1	1	1	1
A4	0	1	1	-	0	1	0	1	1
A5	0	1	1	1	-	1	1	1	0
A6	0	1	1	1	1	-	1	1	0
B1	0	0	0	0	0	0	-	0	0
B2	0	1	1	1	0	1	0	-	1
C1	0	0	0	0	0	0	0	0	-

3.5.4 Creation of Digraph from Antecedent/Reachability sets

Once the final reachability matrix is obtained, the next step is to create the antecedent and reachability sets, which would then help us create a hierarchy of the processes. The process at the top of the resulting digraph has no element in the reachability set except itself, implying that it does not affect any other process and can be removed from the analysis once added to the digraph. Thus, it can be seen that an element whose intersection of reachability set and antecedent set is the reachability set itself can be assigned a position in the digraph and removed from the further steps of the process. In order to break ties, the process with the maximum cardinality of the intersection set is chosen. The series of tables from Table 3.4 to Table 3.9 describe the process involved in obtaining the digraph shown in Figure 3.4

Table 3.4: Partition of processes: Step 1

No.	Antecedent set	Reachability set	Intersection set
A1	A1	A1,A2,A3,A4,A5,A6,B2	A1
A2	A1,A2,A3,A4,A5,A6,B2	A2,A3,A5,A6	A2,A3,A5,A6
A3	A1,A2,A3,A4,A5,A6,B2	A2,A3,A4,A5,A6,B1,B2,C1	A2,A3,A4,A5,A6,B2
A4	A1,A3,A4,A5,A6,B2	A2,A3,A4,A6,B2,C1	A3,A4,A6,B2
A5	A1,A2,A3,A5,A6	A2,A3,A4,A5,A6,B1,B2	A2,A3,A5,A6
A6	A1,A2,A3,A4,A5,A6,B2	A2,A3,A4,A5,A6,B1,B2	A2,A3,A4,A5,A6,B2
B1	A3,A5,A6,B1	B1	B1
B2	A1,A3,A4,A5,A6,B1	A2,A3,A4,A6,B2,C1	A3,A4,A6,B2
C1	A3,A4,B2,C1	C1	C1

Process C1 is removed from the analysis and appended to the digraph

Table 3.5: Partition of processes: Step 2

No.	Antecedent set	Reachability set	Intersection set
A1	A1	A1,A2,A3,A4,A5,A6,B2	A1
A2	A1,A2,A3,A4,A5,A6,B2	A2,A3,A5,A6	A2,A3,A5,A6
A3	A1,A2,A3,A4,A5,A6,B2	A2,A3,A4,A5,A6,B1,B2	A2,A3,A4,A5,A6,B2
A4	A1,A3,A4,A5,A6,B2	A2,A3,A4,A6,B2	A3,A4,A6,B2
A5	A1,A2,A3,A5,A6	A2,A3,A4,A5,A6,B1,B2	A2,A3,A5,A6
A6	A1,A2,A3,A4,A5,A6,B2	A2,A3,A4,A5,A6,B1,B2	A2,A3,A4,A5,A6,B2
B1	A3,A5,A6,B1	B1	B1
B2	A1,A3,A4,A5,A6,B1	A2,A3,A4,A6,B2	A3,A4,A6,B2

Process B1 is removed from the analysis and appended to the digraph

Table 3.6: Partition of processes: Step 3

No.	Antecedent set	Reachability set	Intersection set
A1	A1	A1,A2,A3,A4,A5,A6,B2	A1
A2	A1,A2,A3,A4,A5,A6,B2	A2,A3,A5,A6	A2,A3,A5,A6
A3	A1,A2,A3,A4,A5,A6,B2	A2,A3,A4,A5,A6,B2	A2,A3,A4,A5,A6,B2
A4	A1,A3,A4,A5,A6,B2	A2,A3,A4,A6,B2	A3,A4,A6,B2
A5	A1,A2,A3,A5,A6	A2,A3,A4,A5,A6,B2	A2,A3,A5,A6
A6	A1,A2,A3,A4,A5,A6,B2	A2,A3,A4,A5,A6,B2	A2,A3,A4,A5,A6,B2
B2	A1,A3,A4,A5,A6	A2,A3,A4,A6,B2	A3,A4,A6,B2

Process A2 is removed from the analysis and appended to the digraph

Table 3.7: Partition of processes: Step 4

No.	Antecedent set	Reachability set	Intersection set
A1	A1	A1,A3,A4,A5,A6,B2	A1
A3	A1,A3,A4,A5,A6,B2	A3,A4,A5,A6,B2	A3,A4,A5,A6,B2
A4	A1,A3,A4,A5,A6,B2	A3,A4,A6,B2	A3,A4,A6,B2
A5	A1,A3,A5,A6	A3,A4,A5,A6,B2	A3,A5,A6
A6	A1,A3,A4,A5,A6,B2	A3,A4,A5,A6,B2	A3,A4,A5,A6,B2
B2	A1,A3,A4,A5,A6	A3,A4,A6,B2	A3,A4,A6,B2

Processes A4 and B2 are removed from the analysis and appended to the digraph

Table 3.8: Partition of processes: Step 5

No.	Antecedent set	Reachability set	Intersection set
A1	A1	A1,A3,A5,A6	A1
A3	A1,A3,A5,A6	A3,A5,A6	A3,A5,A6
A5	A1,A3,A5,A6	A3,A5,A6	A3,A5,A6
A6	A1,A3,A5,A6	A3,A5,A6	A3,A5,A6

Processes A3, A5 and A6 are removed from the analysis and appended to the digraph

Table 3.9: Partition of processes: Step 6

No.	Antecedent set	Reachability set	Intersection set
A1	A1	A1	A1

In the end, Process A1 is added to the bottom of the digraph.

This provides us with a digraph that illustrates the hierarchy of the processes involved in our model. This process is important because creating such a hierarchy eases the implementation of an agent based model by allowing us to define the dependencies for each of the agents. Figure 3.4 shows the final digraph formed by the ISM methodology.

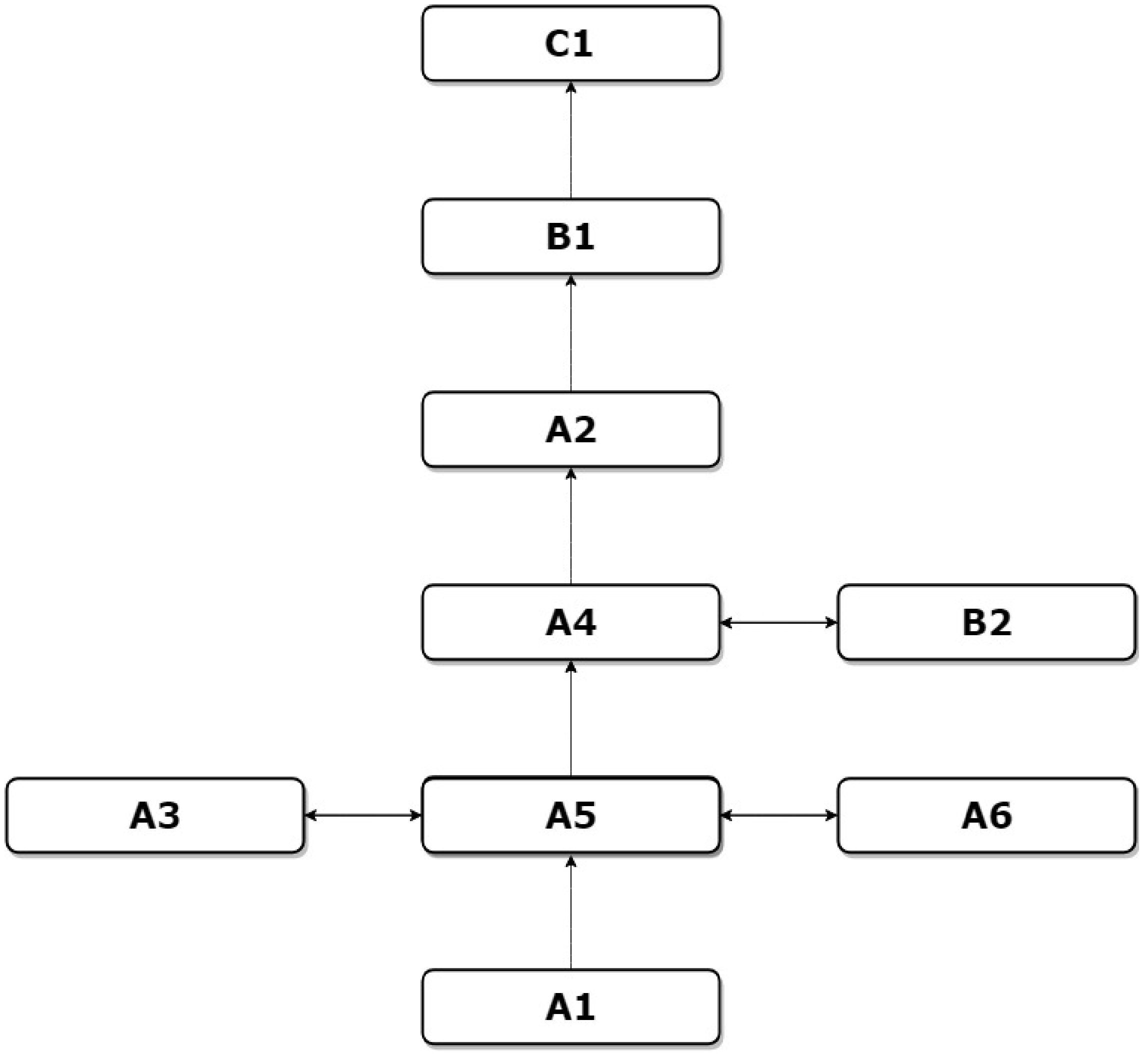

Figure 3.4: The final digraph illustrating the hierarchy of the processes involved

In order to analyze the drive power and dependence of the processes, we carry out a MICMAC *(Matrice d' Impacts Croisés-Multiplication Appliquée á un Classement)* analysis. The analysis is carried out twice, once for the initial reachability matrix (Figure 3.5), and once for the final reachability matrix (Figure 3.6). Using this analysis, we can classify the processes as

- Driver (Strong drive power, Weak dependence)
- Linkage (Strong drive power, Strong dependence)
- Autonomous (Weak drive power, Weak dependence)
- Dependent (Weak drive power, Strong dependence)

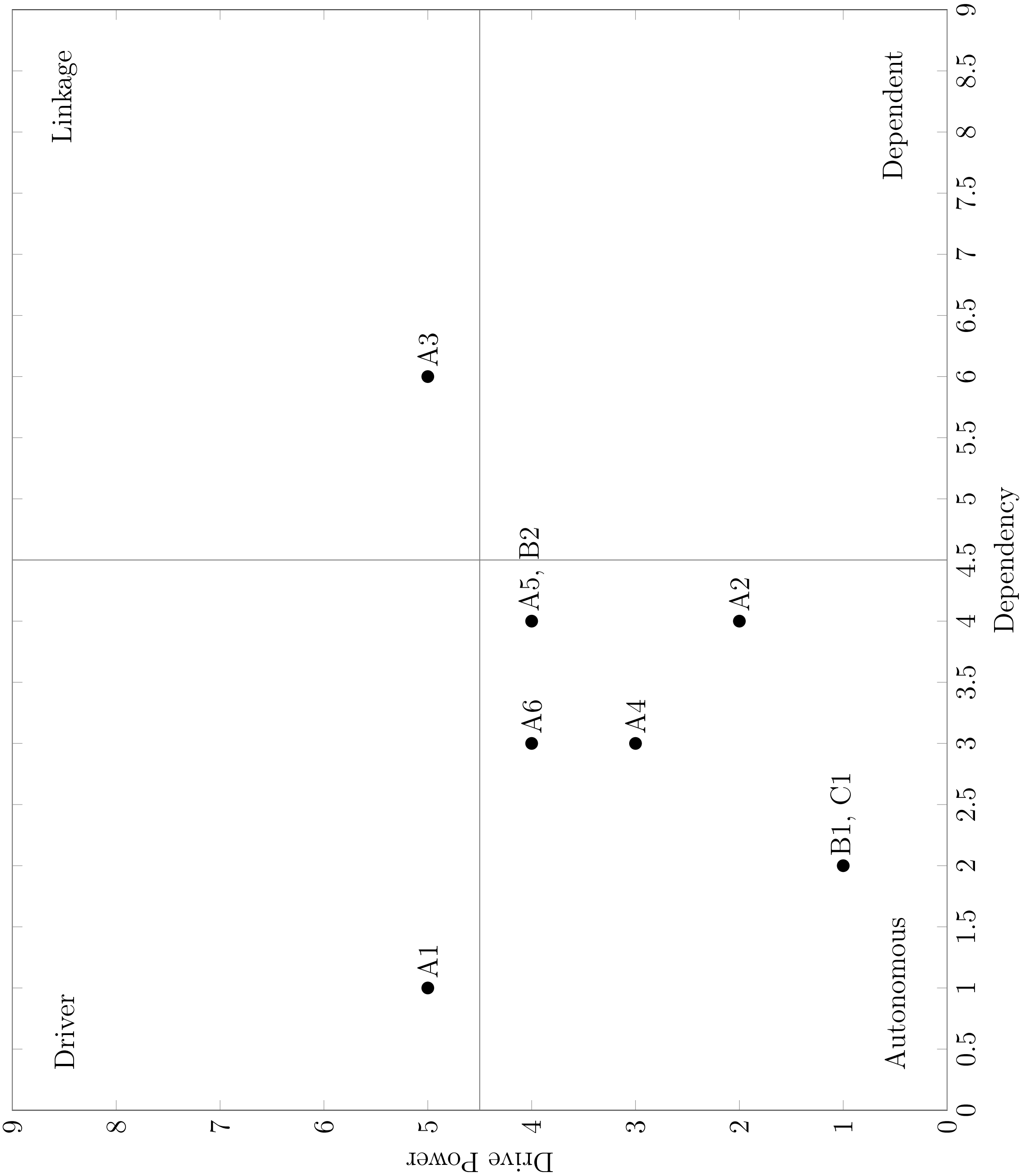

Figure 3.5: MICMAC Analysis for the Initial Reachability Matrix

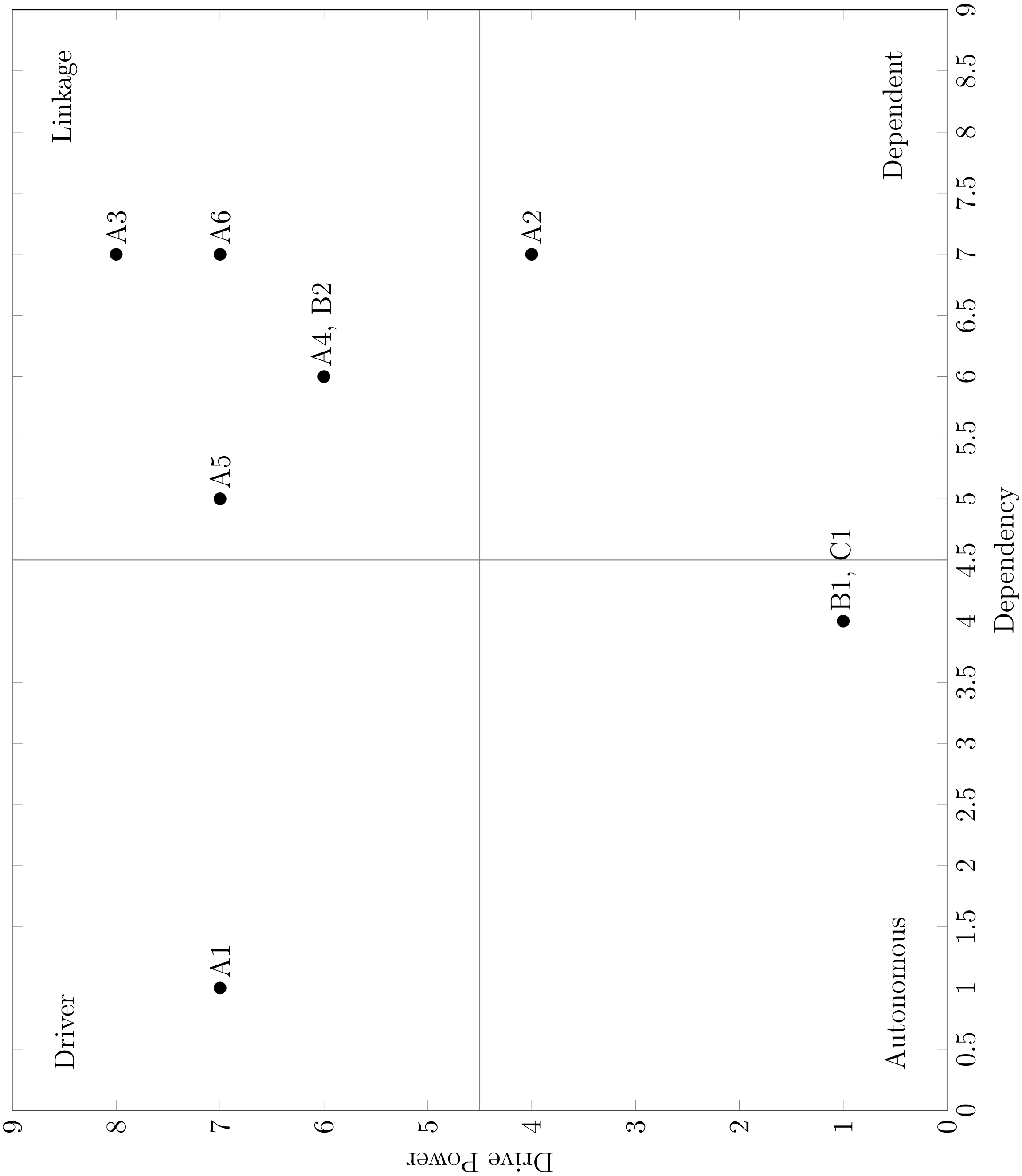

Figure 3.6: MICMAC Analysis for the Final Reachability Matrix

3.6 Conclusion

In this chapter, we propose a novel model for managing first mile reverse logistics for supply chain in online marketplaces. We present a mobile mother-daughter collection system to reduce operational costs and intend to increase costumer satisfaction by bettering the service level for returns. In order to plan the scheduling and route planning logic for the collectors and accumulators, we propose to use multi-agent reinforcement learning algorithms to introduce cooperation.

The application of the ISM methodology and MICMAC analysis provide an important input as to how our processes and subprocesses are linked. The ISM analysis illustrates how the processes are ordered in a hierarchical fashion and helps us understand the properties of each of the agents to be incorporated in the model. The next chapter discusses the work to be done in the second stage.

Chapter 4

Future Work and Plan

The next step is to develop a reinforcement learning algorithm in order to teach the collector agent to locate itself and schedule its collection route optimally. In the first model of our system, the collector agent learns how to approach the collector when required, locate itself in legal parking areas and deposit its collected stock at the accumulator at an optimal frequency that is not too high so as to increase operational costs but also not too low so as to cause an overflow.

A basic sketch of the environment has been shown in Figure 4.1. Customers come up stochastically on the grid with the restriction that only up to one customer may be present in the environment. The accumulator has been kept fixed for now, while the legal parking zone is the left bottom corner from where the collector begins.

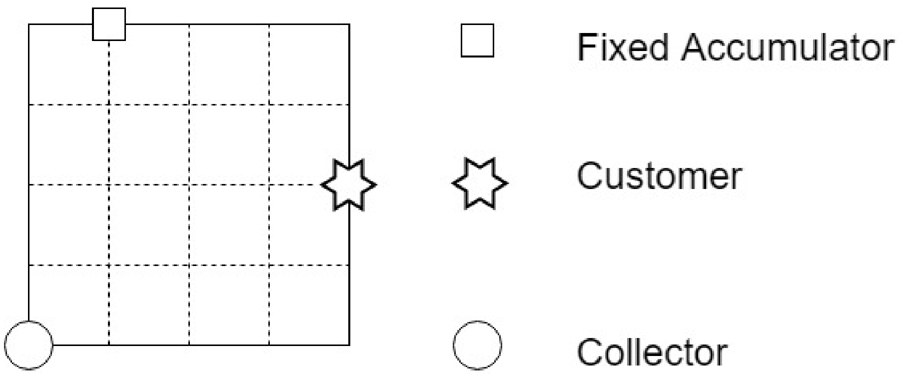

Figure 4.1: Learning Algorithm - Step 1

Once the collector has learnt how to cater to the single customer, the next task for the reinforcement learning algorithm is to find the best route to cater to multiple customers with a fixed accumulator. Customers now appear stochastically on the grid with more than one customers allowed at the same time. Figure 4.2 illustrates a basic sketch of the system in this step.

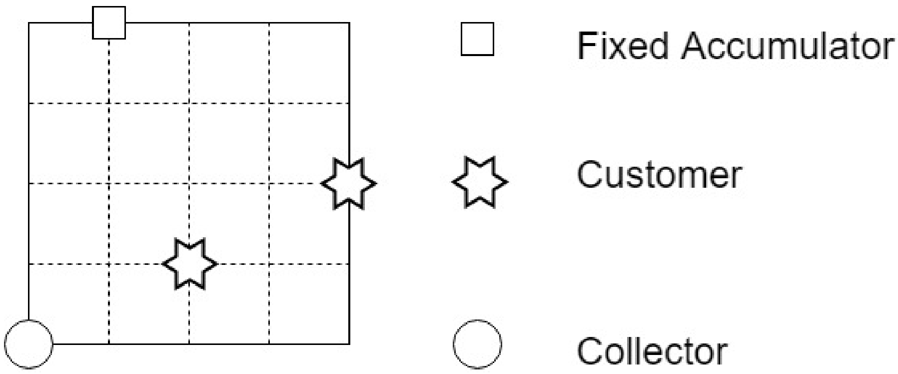

Figure 4.2: Learning Algorithm - Step 2

The next step is the most important and the most difficult part of the process. The task now is for multiple collectors to cooperate and coordinate, and cater to the customers present on the grid. The accumulator in this step is mobile and needs to locate itself optimally in a legal parking zone. Figure 4.3 shows a sketch of the system in this step.

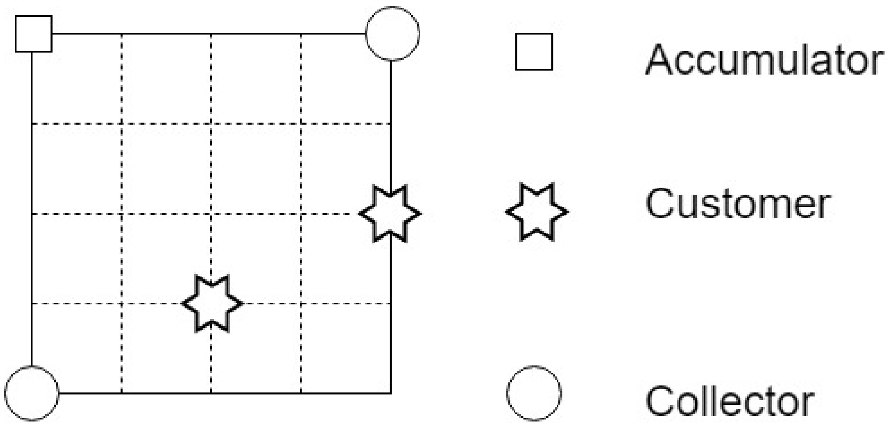

Figure 4.3: Learning Algorithm - Step 3

The results of this analysis would then be compared with conventional methods.

Bibliography

[Agrawal et al., 2015] Agrawal, S., Singh, R. K., and Murtaza, Q. (2015). A literature review and perspectives in reverse logistics. *Resources, Conservation and Recycling*, 97:76–92.

[Akçalı et al., 2009] Akçalı, E., Çetinkaya, S., and Üster, H. (2009). Network design for reverse and closed-loop supply chains: An annotated bibliography of models and solution approaches. *Networks*, 53(3):231–248.

[Akchurina, 2010] Akchurina, N. (2010). *Multi-agent reinforcement learning algorithms.* PhD thesis, University of Paderborn.

[Akdoğan and Coşkun, 2012] Akdoğan, M. Ş. and Coşkun, A. (2012). Drivers of reverse logistics activities: an empirical investigation. *Procedia-Social and Behavioral Sciences*, 58:1640–1649.

[Aras et al., 2008] Aras, N., Aksen, D., and Tanuğur, A. G. (2008). Locating collection centers for incentive-dependent returns under a pick-up policy with capacitated vehicles. *European Journal of Operational Research*, 191(3):1223–1240.

[Attri et al., 2013] Attri, R., Dev, N., and Sharma, V. (2013). Interpretive structural modelling (ism) approach: an overview. *Research Journal of Management Sciences*, 2(2):3–8.

[Avci and Selim, 2016] Avci, M. G. and Selim, H. (2016). A multi-agent system model for supply chains with lateral preventive transshipments: Application in a multi-national automotive supply chain. *Computers in Industry*, 82:28–39.

[Bakker et al., 2010] Bakker, B., Whiteson, S., Kester, L., and Groen, F. C. (2010). Traffic light control by multiagent reinforcement learning systems. In *Interactive Collaborative Information Systems*, pages 475–510. Springer.

[Bellman, 1962] Bellman, R. (1962). Dynamic programming treatment of the travelling salesman problem. *Journal of the ACM (JACM)*, 9(1):61–63.

[Bergenti and Poggi, 2000] Bergenti, F. and Poggi, A. (2000). Exploiting uml in the design of multi-agent systems. In *Engineering Societies in the Agents World*, pages 106–113. Springer.

[Bloembergen et al., 2015] Bloembergen, D., Tuyls, K., Hennes, D., and Kaisers, M. (2015). Evolutionary dynamics of multi-agent learning: A survey. *J. Artif. Intell. Res.(JAIR)*, 53:659–697.

[Boyer et al., 2009] Boyer, K. K., Prud'homme, A. M., and Chung, W. (2009). The last mile challenge: evaluating the effects of customer density and delivery window patterns. *Journal of Business Logistics*, 30(1):185–201.

[Chaharsooghi et al., 2008] Chaharsooghi, S. K., Heydari, J., and Zegordi, S. H. (2008). A reinforcement learning model for supply chain ordering management: An application to the beer game. *Decision Support Systems*, 45(4):949–959.

[de Koster Marisa P. de Brito and van de Vendel, 2002] de Koster Marisa P. de Brito, R. B. and van de Vendel, M. A. (2002). Return handling: an exploratory study with nine retailer warehouses. *International Journal of Retail & Distribution Management*, 30(8):407–421.

[Franklin and Graesser, 1996] Franklin, S. and Graesser, A. (1996). Is it an agent, or just a program?: A taxonomy for autonomous agents. In *International Workshop on Agent Theories, Architectures, and Languages*, pages 21–35. Springer.

[Gatti, 2014] Gatti, C. (2014). *Design of experiments for reinforcement learning*. Springer.

[Govindan et al., 2015] Govindan, K., Soleimani, H., and Kannan, D. (2015). Reverse logistics and closed-loop supply chain: A comprehensive review to explore the future. *European Journal of Operational Research*, 240(3):603–626.

[Griffis et al., 2012] Griffis, S. E., Rao, S., Goldsby, T. J., and Niranjan, T. T. (2012). The customer consequences of returns in online retailing: An empirical analysis. *Journal of Operations Management*, 30(4):282–294.

[Guide Jr et al., 2006] Guide Jr, V. D. R., Souza, G. C., Van Wassenhove, L. N., and Blackburn, J. D. (2006). Time value of commercial product returns. *Management Science*, 52(8):1200–1214.

[Harris, 2010] Harris, L. C. (2010). Fraudulent consumer returns: exploiting retailers' return policies. *European Journal of Marketing*, 44(6):730–747.

[Houghton and Beers, 1961] Houghton, R. A. and Beers, H. S. (1961). The world-wide search and retrieval problem - factors involved in the selection of a rescue/retrieval aircraft. In *SAE Technical Paper*. SAE International.

[Jiang and Sheng, 2009] Jiang, C. and Sheng, Z. (2009). Case-based reinforcement learning for dynamic inventory control in a multi-agent supply-chain system. *Expert Systems with Applications*, 36(3):6520–6526.

[Kovalchuk, 2009] Kovalchuk, Y. (2009). *A Multi-agent decision support system for supply chain management.* PhD thesis, University of Essex.

[Little et al., 1963] Little, J. D., Murty, K. G., Sweeney, D. W., and Karel, C. (1963). An algorithm for the traveling salesman problem. *Operations research*, 11(6):972–989.

[Mahadevan and Theocharous, 1998] Mahadevan, S. and Theocharous, G. (1998). Optimizing production manufacturing using reinforcement learning. In *FLAIRS Conference*, pages 372–377.

[Min et al., 2006] Min, H., Ko, H. J., and Ko, C. S. (2006). A genetic algorithm approach to developing the multi-echelon reverse logistics network for product returns. *Omega*, 34(1):56–69.

[Mollenkopf et al., 2011] Mollenkopf, D. A., Frankel, R., and Russo, I. (2011). Creating value through returns management: Exploring the marketing–operations interface. *Journal of Operations Management*, 29(5):391–403.

[Monostori et al., 2006] Monostori, L., Váncza, J., and Kumara, S. R. (2006). Agent-based systems for manufacturing. *CIRP Annals-Manufacturing Technology*, 55(2):697–720.

[Panait and Luke, 2005] Panait, L. and Luke, S. (2005). Cooperative multi-agent learning: The state of the art. *Autonomous agents and multi-agent systems*, 11(3):387–434.

[Rogers and Tibben-Lembke, 2001] Rogers, D. S. and Tibben-Lembke, R. (2001). An examination of reverse logistics practices. *Journal of business logistics*, 22(2):129–148.

[Russell et al., 1995] Russell, S., Norvig, P., and Intelligence, A. (1995). A modern approach. *Artificial Intelligence. Prentice-Hall, Egnlewood Cliffs*, 25:27.

[Sangwan, 2017] Sangwan, K. S. (2017). Key activities, decision variables and performance indicators of reverse logistics. *Procedia CIRP*, 61:257–262.

[Shen et al., 2006] Shen, W., Hao, Q., Yoon, H. J., and Norrie, D. H. (2006). Applications of agent-based systems in intelligent manufacturing: An updated review. *Advanced engineering INFORMATICS*, 20(4):415–431.

[Srinivasan et al., 2002] Srinivasan, S. S., Anderson, R., and Ponnavolu, K. (2002). Customer loyalty in e-commerce: an exploration of its antecedents and consequences. *Journal of retailing*, 78(1):41–50.

[Stock et al., 2006] Stock, J., Speh, T., and Shear, H. (2006). Managing product returns for competitive advantage. *MIT Sloan Management Review*, 48(1):57.

[Stock and Mulki, 2009] Stock, J. R. and Mulki, J. P. (2009). Product returns processing: an examination of practices of manufacturers, wholesalers/distributors, and retailers. *Journal of Business Logistics*, 30(1):33–62.

[Stockheim et al., 2003] Stockheim, T., Schwind, M., and Koenig, W. (2003). A reinforcement learning approach for supply chain management. In *1st European Workshop on Multi-Agent Systems, Oxford, UK*.

[Sutton and Barto, 1998] Sutton, R. S. and Barto, A. G. (1998). *Reinforcement learning: An introduction*, volume 1. MIT press Cambridge.

[Tanaka and Yoshida, 1999] Tanaka, Y. and Yoshida, T. (1999). An application of reinforcement learning to manufacturing scheduling problems. In *Systems, Man, and Cybernetics, 1999. IEEE SMC'99 Conference Proceedings. 1999 IEEE International Conference on*, volume 4, pages 534–539. IEEE.

[Van Hoek, 1999] Van Hoek, R. I. (1999). From reversed logistics to green supply chains. *Supply Chain Management: An International Journal*, 4(3):129–135.

[Veluscek, 2016] Veluscek, M. (2016). *Global supply chain optimization: a machine learning perspective to improve caterpillar's logistics operations*. PhD thesis, Brunel University London.

[Walsh et al., 2014] Walsh, G., Möhring, M., Koot, C., and Schaarschmidt, M. (2014). Preventive product returns management systems-a review and model.

[Wooldridge and Jennings, 1994] Wooldridge, M. and Jennings, N. R. (1994). Agent theories, architectures, and languages: a survey. In *International Workshop on Agent Theories, Architectures, and Languages*, pages 1–39. Springer.

[Wooldridge and Jennings, 1995] Wooldridge, M. and Jennings, N. R. (1995). Intelligent agents: Theory and practice. *The knowledge engineering review*, 10(2):115–152.

[Wu and Dunn, 1995] Wu, H.-J. and Dunn, S. C. (1995). Environmentally responsible logistics systems. *International Journal of Physical Distribution & Logistics Management*, 25(2):20–38.

[Young et al., 2005] Young, L. A., Pisanich, G., Ippolifo, C., and Alena, R. (2005). Aerial vehicle surveys of other planetary atmospheres and surfaces: imaging, remote-sensing, and autonomy technology requirements. Technical report, NATIONAL AERONAUTICS AND SPACE ADMINISTRATION MOFFETT FIELD CA ROTORCRAFT DIVISION.

[Zolfpour-Arokhlo et al., 2014] Zolfpour-Arokhlo, M., Selamat, A., Hashim, S. Z. M., and Afkhami, H. (2014). Modeling of route planning system based on q value-based dynamic programming with multi-agent reinforcement learning algorithms. *Engineering Applications of Artificial Intelligence*, 29:163–177.