Coordinated mobility on demand model for taxi fleet management

Avalpreet Singh Brar

2021

Avalpreet Singh Brar (2021). Coordinated mobility on demand model for taxi fleet management. Master's thesis, Nanyang Technological University, Singapore.
<https://hdl.handle.net/10356/154397>

<https://hdl.handle.net/10356/154397>

<https://doi.org/10.32657/10356/154397>

This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License (CC BY-NC 4.0).

Downloaded on 26 Apr 2023 20:17:01 SGT

**COORDINATED MOBILITY ON DEMAND MODEL FOR
TAXI FLEET MANAGEMENT**

AVALPREET SINGH BRAR

School of Electrical & Electronic Engineering

A thesis submitted to the Nanyang Technological University
in partial fulfilment of the requirement for the degree of
Master of Engineering

2021

Statement of Originality

I hereby certify that the work embodied in this thesis is the result of original research, is free of plagiarised materials, and has not been submitted for a higher degree to any other University or Institution.

7/12/2021

.....
Date

NTU NTU NTU NTU NTU NTU NTU NTU
NTU NTU NTU NTU NTU NTU NTU NTU
Avalpreet Singh Brar
NTU NTU NTU NTU NTU NTU NTU NTU
.....

Avalpreet Singh Brar

Supervisor Declaration Statement

I have reviewed the content and presentation style of this thesis and declare it is free of plagiarism and of sufficient grammatical clarity to be examined. To the best of my knowledge, the research and writing are those of the candidate except as acknowledged in the Author Attribution Statement. I confirm that the investigations were conducted in accord with the ethics policies and integrity standards of Nanyang Technological University and that the research data are presented honestly and without prejudice.

December 7, 2021

.....
Date

.....

.....
Su Rong

Authorship Attribution Statement

This thesis contains material from 2 from papers accepted at conferences in which I am listed as an author.

The contents in the Chapters 2 and 3 appear in A. S. Brar and R. Su, "Dynamic Supply-Demand Balancing Policy for CMoD Fleet," 2021 IEEE 24th International Conference on Intelligent Transportation Systems (ITSC), 2021,

The contributions of the co-authors are as follows:

- A/Prof Su Rong provided the initial project direction and edited the manuscript drafts.
- I formulated the mathematical model and prepared the manuscript drafts.

The contents in Chapter 4 appear in A. S. Brar and R. Su, "Ensuring Service Fairness in Taxi Fleet Management," 2020 IEEE 23rd International Conference on Intelligent Transportation Systems (ITSC), 2020, pp. 1-6, \ doi: 10.1109/ITSC45102.2020.9294504.

The contributions of the co-authors are as follows:

- A/Prof Su Rong provided the initial project direction and edited the manuscript drafts.
- I formulated the mathematical model and prepared the manuscript drafts.

7/12/2021

.....
Date

NTU NTU NTU NTU NTU NTU NTU NTU
NTU NTU NTU NTU NTU NTU NTU NTU
Avalpreet Singh Brar
NTU NTU NTU NTU NTU NTU NTU NTU
.....

AVALPREET SINGH BRAR

Acknowledgements

First, I would like to thank my supervisor, Dr. Su Rong, for giving me this opportunity to pursue my Master of Engineering under his expert guidance. Prof Su Rong provided an optimal environment for learning new skills and implementing them in real-world applications. Second, I would like to thank my parents for their continuous support and guidance. A special thanks to my friends Mridul Marwaha, Parikshit and Harmandeep Singh who have been more like brothers. Above all I would like to thank the lord almighty for this wonderful experience.

Table of Contents

Acknowledgement	i
Summary	iv
Acronyms	v
Symbols	vi
Lists of Figures	viii
Lists of Tables	ix
1 Introduction	1
1.1 Introduction to Mobility-on-Demand (MoD) System	1
1.1.1 MoD system markets	3
1.1.2 Literature Review	4
1.1.3 Motivations and Thesis Organization	18
2 Coordinated Mobility-on-Demand Model	20
2.1 Introduction to CMoD Fleet Management	20
2.1.1 Problem Statement	21
2.1.2 Hierarchy of CMoD System	22
2.1.3 Taxi Operational Network	23
2.1.4 Taxi Trips Data	24
2.1.5 Components of CMoD System	25
2.2 Mathematical Model for CMoD Fleet Management	32
2.2.1 Supply Scheduling	33
2.2.2 Vehicle Rebalancing	36
2.2.3 CMoD fleet control	41
2.2.4 Model Variable Summary	44
2.3 Analysis of the Impact of Vehicle Rebalancing	45
2.4 Conclusion	46
3 Learning Optimal Model Parameters	47
3.1 Learning the Optimal Parameters	47
3.1.1 Motivation	47
3.1.2 Methodology	51

3.1.3 Algorithm	52
3.2 Impact of Supply-Demand Balancing Policy	53
3.2.1 Impact of Implementing Vehicle Rebalancing	53
3.2.2 Case Studies	55
3.3 Conclusion	61
4 Ensuring Service Fairness in CMoD Model	62
4.1 Motivations and Problem Statement	62
4.2 Mathematical Formulation	64
4.2.1 Fleet Management Objectives:	64
4.2.2 Decision Variable	65
4.2.3 Cost Function	66
4.2.4 Constraints:	69
4.2.5 Multi-Objective Taxi Fleet Management Problem Summary	71
4.3 A Receding Horizon Control Strategy	72
4.4 Simulation Case Studies	74
4.4.1 Company Benefits	75
4.4.2 Driver Benefits	77
4.5 Conclusion	81
5 Summary and Possible Future Directions	82
5.1 Summary	82
5.2 Future Directions	84
Publications	85
Bibliography	86

Summary

In the current mainstream industrial practice, Mobility-on-Demand (MoD) systems work in a partially coordinated manner. A centralized driver-allocation system matches a passenger request to the closest idle driver, and the drivers typically follow their own passenger finding strategies. This uncoordinated behavior among drivers leads to a supply-demand imbalance over the operating area. Fleet coordination is a potential solution to mitigate the supply-demand imbalance. In this dissertation, we present our Coordinated Mobility-on-Demand (CMoD) system model for optimal vehicle scheduling and rebalancing. We provide a detailed analysis of the model parameters such as the balancing frequency and the prediction horizon, that are critical for its real-time implementation and assess the performance of the proposed model in a simulated environment. Finally, we present an extension of our model to ensure service fairness for the drivers while operating in this coordinated environment. We quantify service fairness in terms of low disparity of drivers' incomes and high booking success rate and show that the proposed extension of our model leads to a significant improvement in service fairness.

Keywords: CMoD, Supply-Demand Balancing, Service Fairness

Acronyms

- CMoD - Coordinated Mobility on Demand System
- AMoD - Autonomous Mobility on Demand System
- MILP - Mixed Integer Linear Program
- WGC - Wild Goose Chase
- RHC - Receding Horizon Control
- UCB - Upper Confidence Bound
- TLC - Taxi and Limousine Commission
- SDR - Served Demand Ratio
- MDP - Markov Decision Process
- NYC - New York City
- TRC - Total Recommendation Count
- SRC - Scheduled Recommendation Count
- SGR - Scheduled Guarantee Ratio
- EoD - End of Day
- EoM - End of Month
- KPI - Key Performance Indicator

Symbols

- μ - per mile charge in dollars/mile
- ξ - fuel cost in dollars/gallon
- ζ - fixed charge in dollars
- χ - mileage in gallons/mile
- f - balancing frequency
- ω - prediction horizon
- η - recommendation frequency
- κ - fraction of company's share
- φ - reachability index

List of Figures

1.1 Mobility on Demand System	2
1.2 Demand curve [7]	3
1.3 Supply-Demand Equilibrium [7]	3
1.4 Temporal evolution	4
1.5 Spatial Distribution	4
1.6 Vehicle Rebalancing	8
1.7 Recommender System	12
1.8 Dynamic Pricing	15
2.1 Elements of CMoD Model	21
2.2 CMoD Hierarchy	22
2.3 Manhattan Taxi Zones	23
2.4 Yellow Taxi Trip Data	24
2.5 Supply Matrix	25
2.6 Noise Analysis	28
2.7 Anticipated Supply and Demand	29
2.8 Source: NYC Taxi and Limousine Commission [8]	31
2.9 Coordinated Mobility-on-Demand Objectives	32
2.10 Supply Scheduling	33
2.11 Vehicle Rebalancing	37
2.12 Impact of Vehicle Rebalancing	46
3.1 Frequency of Vehicle Rebalancing	48
3.2 Horizon Window	50
3.3 Single Step Markov Decision Process	51
3.4 Passenger Demand Pattern (Manhattan Jan, 20 2020)	53
3.5 Learning Optimal Set of Parameters	54
3.6 Horizon Window Analysis	55
3.7 Frequency Analysis	57
3.8 Demand Market Analysis	58
3.9 Impact of Demand Prediction Inaccuracies	60
4.1 Service Fairness	63
4.2 Multi-Objectives of Taxi Fleet Management	65
4.3 Multi-Objective Taxi Fleet Management Problem	71
4.4 Coordinated Mobility-on-Demand System Fleet Control	73
4.5 Daily served demand wrt actual demand for 100 Taxis	75

4.6	Daily served demand wrt actual demand for 100 Taxis	76
4.7	Profit earned by drivers at month end for different Models	77
4.8	Profit disparity among drivers over a period of 30 days	78
4.9	Scheduled Recommendation Count Analysis	79

List of Tables

2.1 CMoD Model Variables	44
------------------------------------	----

Chapter 1

Introduction

1.1 Introduction to Mobility-on-Demand (MoD) System

A Mobility-on-Demand system in the form of taxi system which involves pickup and drop-off service has existed for many years now. However, the ride sharing systems have also significantly grown their market share [17]. These ride-sharing platforms such as Uber and Grab have effectively demonstrated the potential for road vehicles to be used as a sustainable mode of transportation in the rapidly expanding urban environments. The introduction of e-hailing, i.e., booking a ride online in the MoD system has opened up the opportunity to match the taxi drivers to passenger requests more efficiently. In comparison to the street-hailing taxi systems, these platforms have significant advantages. Firstly, they allow multiple passengers (via carpooling) to share a single vehicle which increases the capacity utilization of the taxi fleet and also higher level of demands can be served. Secondly, the vehicle routes of the taxis can be optimized this reduces the travel time, hence improving the quality-of-service of the system. [5].

Another salient feature of the MoD system is that the taxis provide a door-to-door pick-up and drop-off service, i.e., there are no fixed boarding and alighting stops like in the case of the bus and the metro transport systems. Passengers make up pickup requests with a predefined origin-destination pair. A centralized passenger-taxi matching platform allocates an idle standing taxi to a passenger request. The allocated taxi driver drives to the pick-up location, the passenger boards the taxi and the driver drop-offs him at the specified destination. The MoD systems provide a certain degree of freedom, the passengers can cancel

the pick-up requests if the waiting time or the trip price is too high for them.

Fig 1.1 explains a simplistic view of the current practices in the MoD system operations. The passengers make a pick-up request, the driver allocation platform finds the nearest driver and matches with the passenger to complete the trip by first picking up the passenger from his location and then dropping-off him to his destination. From a passenger's point of view, low waiting time and low trip rates are the key performance indicators of the MoD platform's quality of service. However, for the drivers, the low idle mileage to pick up the passenger, and between two subsequent rides which influences their profits are the major concerns.

Figure 1.1: Mobility on Demand System

At the fleet level, the operational efficiency of the MoD system, can be assessed from factors such as the number of taxis in the system, the percentage utilization of taxis, the number of passenger requests that were not served, and the average passenger waiting time [4]. The authors in [17] rightfully argue that neither high wait times alone nor low taxi utilization rates can be interpreted as a sign of inefficiency. In contrast, a simultaneous occurrence of empty taxis (low utilization) and long waiting times (low service level) indicates an inefficient taxi allocation. Therefore, if wait times are unusually high for a given level of occupancy, or if occupancy is unusually low for a given level of wait times, the system performance is suffering from an inefficient assignment of demand and supply.

1.1.1 MoD system markets

In MoD system setting, the served taxi trips are treated as the quantity demanded. The demand curve $D(T, p)$ is a function of pick-up time T and ride-fare p . The quantity demanded Q increases with a decrease in the waiting time as well as the ride-fare. Fig 1.2 show the demand curve which was taken from the analysis done by [7]. This demand curve is only representational and haven't been used to generate demand. In this thesis, we use the real demand data [8].

Figure 1.2: Demand curve [7]

The quantity supplied, i.e., the served passenger requests depend on the availability of the idle taxis. The supply curve $S(T, p)$ increases with an increase in ride-fare p as well as the average waiting time T . Authors in [7] present an interesting phenomena. The supply curve bends backward for higher waiting times. This situation occurs when the taxi supply is very thinly distributed over the operating area and drivers start dropping out due to large pick-up times. This operational area shown in red is called the wild goose chase region. The operating point must not lie in this region. It is always desirable that the operating point (equilibrium) lies in the green region as explained in detail in [7].

Figure 1.3: Supply-Demand Equilibrium [7]

1.1.2 Literature Review

Passenger Demand Prediction

The demand distribution varies spatially and temporally, i.e., the demand in some regions is higher than the other regions during a particular time duration and the demand varies in a particular region over time as well.

Temporal evolution of demand

Fig 1.4 shows the temporal evolution of the overall passenger pick-up demand throughout the day. This data was extracted using the taxi trips data [8] for Manhattan city. As seen in the data there are two peak hours in a day corresponding to the office hours. The demand varies over time during the day.

Figure 1.4: Temporal evolution

Spatial distribution of demand

Fig 1.5 shows the spatial distribution of the passenger demand. City of Manhattan is divided into 66 regions. The following figure shows, the number of passenger requests (y-axis) vs region id (x-axis). As seen in the figure, some regions are inherently more popular than the other regions and the pick-up and drop-off demand for these regions are higher than the other regions.

Figure 1.5: Spatial Distribution

The authors in [20] presented a model to predict the passenger request count at each taxi stand. The proposed model predicts the passenger demand over the operating area for a short horizon window. This model utilizes the stream of data that is transmitted by the telematics that are installed in each taxi for the present time slot in-order to predict the future outcome. The contribution of this paper is to predict at the instant (t), the number of pick-up request that will come out during the horizon $[t, t + P]$ at each taxi stand. The process is repeated for the next time duration, i.e., for the following instant ($[t + P, t + 2P]$) by reusing the actual value of the passenger requests during ($[t, t + P]$), and so on (i.e., the framework continuously runs in a stream). The proposed framework was tested online using a real taxi fleet. For the case study, a taxi fleet from the city of Porto, Portugal was selected. Porto has a total of 63 taxi stands and there are two competing taxi companies, that have their own taxi fleet. The streaming transmitted by the larger company having 441 vehicles was used for this analysis. The average idle time, i.e. the time that each taxi driver has to wait for his next passenger was found out to be 44 minutes. The authors used time-series forecasting techniques such as the time-varying Poisson model and the auto regressive integrated moving average (ARIMA) for predicting the passenger demands. The novelty of this paper was to build a model for predicting the spatial temporal distribution of the passenger pickup demand using streaming data.

The operational efficiency, in terms of higher occupancy rates as well as the quality of service in terms of lower passenger waiting time can be achieved by predicting the passenger demands. In [19], the authors proposed a machine learning based model that uses the current and historical passenger demands as features, in-order to predict the passenger demands in each region of the city. The hypothesis of the authors is that a detailed past information is critical for accurately determining the passenger demands in the future, since the pick-up requests in the near future are correlated with the events that happened in the past. The authors use an example of someone who requests a taxi to a shopping center, may also request a taxi to return home after few hours. In order to capture these long term dependencies, the authors use long short term memory (LSTM) sequence model that has a gating mechanism that stores the past information in a hidden state, in-order to predict the future information. The authors test their model using the data set of passenger pick-up requests in city of New York by dividing the city into small zones and predicting the demand in each zone. The authors claim that the proposed approach outperforms other prediction methods, such as the feed-forward neural networks.

Taxi Fleet Coordination

Fleet coordination means that a central scheduler shall decide the operational policies for the taxi fleet. This means that the scheduler shall recommend the idle standing drivers a suitable location to wait for their next ride. The main advantage of this strategy is that better coordination leads to mitigation of supply-demand imbalance and improves the quality of service such as the low waiting times and low request drop-outs. Secondly, due to higher served demands the driver as well as the company profits are expected to grow. Various authors have proposed fleet coordination as a promising upgrade to the contemporary MoD system.

Authors in [26] proposed three strategies that differ by the amount of information exchanged between the drivers and the passengers and the degree of the cooperation between taxi drivers and the central coordinator. The authors used a high fidelity simulation environment (MATSim) to evaluate three strategies. On top of this, the authors used a dynamic vehicle routing optimizer (DVRP Optimizer). The performance evaluation was done for the city of Mielec. The authors explain that the cooperation between the dispatcher and the taxi drivers is of critical importance. However, the customer-dispatcher communication may be substituted by the use of more sophisticated dispatching strategies, thereby not affecting the quality of service.

Authors in [17] show that fleet coordination can improve the operational efficiency and the quality of service of the existing taxi systems. Depending on the taxi market, i.e. the city of operation, the idle vehicle miles and the taxi fleet sizes could be reduced in a significant manner. Also, the passenger waiting time can be reduced which is quantified as an improvement in the quality of service, for the same fleet size. The authors propose that full fleet coordination, which is generally considered feasible only for the fully autonomous systems, can also be implemented in the existing taxi systems as well. The authors have quantified the benefits of implementing the fleet coordination in the existing MoD systems using simulation based studies for the cities of San Francisco, Chicago and Zurich.

The tutorial paper [25] examines the various aspects of an Autonomous Mobility-on-Demand (AMoD) system that operates in a coordinated manner. Coordinated AMoD system involves robotic, self driving vehicles that provide the pick-up and drop-off services to the passengers. The authors highlighted three dimen-

sions of the AMoD system: (1) models that can facilitate in capturing the dynamics and the stochastic aspects of passenger demands (2) fleet coordination algorithm that maximizes the operational efficiency and the quality of service (3) economic aspects – cost to the company, impacts of the markets, etc.. This work was aimed at gauging the value of the AMoD system. The case study of New York Taxi data showed that same level of demands can be served using 70% of the current fleet size whereas for Singapore, the entire countries mobility demands can be served by using a fleet of AMoD vehicles that have a size of 40% of the current fleet size.

The authors in [18] analyzed the performance of four different operational policies to control an AMoD system. The authors used a high fidelity simulation environment to replicate the real-world environment and implemented a sequential vehicle-sharing model. The authors used a captured the road level traffic dynamics while performing the simulation, and the study was conducted for the Zurich city. The simulation results showed that although the cost of the AMoD systems was considerably higher than the public transport system or the private cars, the system showed a significant improvement on the capacity utilization. It was observed that the proposed model led to six times higher occupancy rates. The authors indicate that the justification of using AMoD system becomes clear only in long-term utilization. Simulations demonstrated the significance of the fleet sizing and how the operational policies such as the vehicle rebalancing lead to a significant improvement on the waiting time reduction with a minor increase in the cost of operation.

Previous works have shown that with the implementation of the fleet coordination, it would be possible to have a better control over the supply distribution. Many operational strategies have been proposed to mitigate the supply-demand imbalance. There are primarily three strategies in the literature to deal with the supply-demand imbalance problem namely vehicle rebalancing systems, recommender systems and dynamic pricing.

Vehicle Rebalancing Systems

The central idea behind the vehicle rebalancing is that if we could anticipate the expected value of the passenger pick-up requests in every region of the operating area, we can then dispatch the idle standing taxis from one region to the other in advance, so that higher demands can be served in the near future. The figure below, shows the city of Manhattan. Assuming that we can anticipate the demands correctly, the taxis can be dispatched towards the anticipated demands in advance. The entire operating area is divided into regions, also known as the taxi zones. Using vehicle rebalancing, the goal is to reduce the supply and demand imbalance in all the regions in the near future by re-distributing the idle standing taxis from one region to the other.

Figure 1.6: Vehicle Rebalancing

Authors in [3] explain that due to the disparity in the popularity of various pick-up and drop-off points, it is highly likely that the taxis start gathering at the popular locations and are depleted at the other regions. Such a scenario can cause an imbalance in the supply-demand distribution. The proposed solution is to re-distribute the idle standing vehicles by predicting the passenger demands. This makes vehicle rebalancing strategy critical for improving the operational efficiency and quality of service of the mobility-on-demand system. The authors developed a model for maximizing the throughput of a MoD system while operating in an urban environment. They consider a fleet of taxis that wait for the passengers to come at a set of taxi stations. Passengers arrive at the taxi stations to pick-up the vehicle and drive it to the destination station, where they drop-off the vehicle. The authors have proposed a robotic solution to rebalance the idle standing robotic vehicles that can autonomously drive between stations. A fluid model was used by the authors for the passengers and the taxis in the system. A supply rebalancing policy was developed that can ensure that every station can reach an equilibrium point, i.e., an excessive built up of the passengers does not take place at the stations. Also, the objective is to ensure that the proposed rebalancing solution leads to a minimum cost of operation.

Authors in [2] developed a mobility-on-demand system model which acted as a framework for taxi dispatching. The objective of this model was to match the spatio-temporal ratio between the demand and supply for service quality. The authors also ensured that this objective is met while the taxi idle driving distance is minimized. They used the data-set containing the trip records from San Francisco, to simulate and test their model. The proposed model showed that the proposed vehicle rebalancing solution reduces the gap in the supply and demand as well as leads to lower average total idle distance. In [1], the authors presented their work to make their formulation robust with respect to the demand uncertainties.

Authors in [11] proposed a reinforcement learning to perform the vehicle rebalancing strategy. More specifically, the authors use a Q-Network to achieve this goal. This is a model free approach, unlike the network based approaches as mentioned above, that require explicit modelling of every detail in the vehicle rebalancing system. They apply this approach to an agent based simulator and test it on a London case study. Results show that, the proposed method can perform better than the optimization model based methods that first predict the passenger demands and then generate a dispatch solution to maximize the rewards.

Authors in [12] also used a reinforcement learning approach for performing the vehicle rebalancing in the MoD system. Their approach was to use a batch training algorithm that learns the driver-perspective state-value function using the spatial-temporal deep value networks. Later, value-based policy search was used to determine the optimal re-positioning action. This approach combines decision-time planning and bootstrapping with the value networks. In-order to test the efficacy of the proposed method, the authors used the baseline in the existing MoD system. It was observed that the proposed solution led to an improvement in the income rate. Authors also performed an experiment involving the regular drivers on their own MoD platform to test it's performance in the real-world. The authors compared the performance of the drivers performing the vehicle re-positioning using their model with the drivers who used their own experience to find their next ride.

Authors in [6] presented a mathematical model that also utilizes car-pooling strategy along with vehicle rebalancing. The authors claim that their model can be scaled to a large number of passengers and can handle huge number of trips. Apart from this, they can also dynamically generate the optimal routes w.r.t the spatial-temporal distribution of the supply and demand. The authors quantified experimentally the impact for fleet size and capacity of vehicles (low and medium capacity) on the operational efficiency and quality of service that is quantified by the average waiting time, travel delays, etc,. The algorithm was validated using a large scale dataset from the city of New York. Their study considered that the vehicle capacity can be as large as 10 passengers per vehicle.

Authors in [23] proposed the operational policies for the N-Vehicle Dynamic Traveling Repairman Problem and the N-Vehicle Dynamic Vehicle Routing Problem. This approach takes into account the fact that the historical demand distribution can be used to predict the demands in the future. This learned information regarding the future demand helps the system to plan which robot has to be sent to which location. The proposed policies are model-free and the authors provide performance bounds for their proposed solution for the time-invariant setting where the arrival of the demands is at a constant rate, both spatially and temporally. This work is then extended to a time-varying case, and test the efficacy of the proposed methodology. Authors apply the proposed operational policies to MoD fleet operation in simulation using AMoDeus.

Authors in [24], investigate the autonomous vehicle sharing and reservation (AVSR). In such a setting, the passenger can request for an Autonomous Vehicle for a trip ahead of time, at which he intends to start the trip and the reservation system operator will arrange the AV for picking up the passenger and drop him to his destination. The goal is to design a chain of trips based on the trip requests that have been recorded so far. Instead of using the vehicle rebalancing strategy, the authors propose that a sequence of potential trips can be determined using the historical data which can serve the maximum demands over a long period of time.

Authors in [27] presented an approach to optimize the taxi ride matching and routing for a fleet of Autonomous Vehicles. The proposed discrete-time model has the potential for integrating the practical operational constraints such as the charging constraint into the the system. They use model predictive control (MPC) to reduce the computational complexity. Secondly, leveraging their model, and the model predictive control algorithm the authors seek an optimal coordination of an AMoD system. They prove the system stability in the sense of Lyapunov. The objective of this work is to provide a decision on whether the taxi should wait for the customer to arrive at the station that the taxi is already standing, or to serve a current demand and also if the idle standing vehicle should go and wait at the other location for his next ride.

The authors in [28] used a queuing-theoretic framework to provide a solution to the vehicle rebalancing problem for the MoD systems. Like other previous works, this approach serves as a tool for making the dispatch decisions in a coordinated fleet setting.

Recommender Systems

Apart from a fleet level coordination system like the vehicle rebalancing system, some authors have also proposed the recommender systems for individual drivers. A recommender system suggests a suitable location for the idle standing driver that maximises his chance to get a ride. The primary difference between a recommender system and the vehicle rebalancing system is that the recommender system generally don't take the view point of maximizing the company level benefits. Rather, the objective is to maximize the chance of the idle standing driver to get the ride.

Figure 1.7: Recommender System

According to the authors in [33], taxi drivers need efficient strategies to quickly find their next rides so that their daily earnings can be maximized. It has been observed that there is a huge disparity in the passenger finding capabilities of different drivers in the fleet. This is primarily because different drivers have different experience level. The top performing taxi drivers can earn 25% more than the novice drivers, even though both the drivers are operating in the same taxi market and using the same ride matching technology. A better strategy has dual benefits. Firstly, it reduces the driver's idle waiting time. Secondly, it leads to lower level of carbon emissions. Providing suitable recommendations to the drivers who are operating in a stochastic demand market is a challenging task. Recently, due to the availability of massive amounts of data pertaining to the taxi trip records, new frontiers have opened up with regards to improving the system recommendation capabilities. The objective of this paper is to mine the efficient operational strategies for the drivers so that their profits can be maximized. The performance the operational strategies is assessed against the metrics, such as the taxi trip frequency, the hot spots of passenger demands, and the idle taxi mileage. The authors used the techniques such as the Newton's polynomial interpolation and Gamma distribution, to understand their dynamics. Their strategy is to increase the taxi driver income levels by recommending them a route through the demand hot spot, which is determined using the actual taxicab traces from the city of Changsha.

Authors in [13] presented a dual recommender system that gives the recommendations to the taxi drivers as well as the people expecting to take a taxi. In order to achieve this, the authors used the pattern of the passenger demand. Apart from this, the GPS trajectories of the taxi drivers' picking-up/dropping-off points was also used. First, the system provides the drivers, a point-of-interest, i.e., a specific (latitude, longitude) point to wait for their next ride. The system also recommends a route to that point that maximizes their probability of getting a request while travelling. Second, the system provides a specific point for the rider to wait for quickly getting a driver. In their study, the authors use the historical GPS trajectories of taxis to provide the recommendations to the riders and the drivers. They feed the learned information into a probabilistic model which estimates the profit that can be earned by the candidate driver based on where and when the driver asks for the recommendation.

Authors in [14] developed a recommender system for taxi drivers with the objective to maximize their earnings when the driver follows the route recommended by the system to find the passenger. Specifically, they first define a

profit based objective function. Using the historical taxi GPS traces, the authors developed a graph representation of road networks. To generate optimal driving route for recommendation, the authors use a brute-force strategy. Particularly, the system is capable of providing an entire driving route instead of recommending a sequence of pick-up points and letting the driver decide how to get to those points.

Authors in [32] used the GPS traces to study of the energy-efficient transportation patterns. Specifically, the proposed system provides a sequence of mobile recommendations. A mobile recommender system was developed that had an ability of recommending a sequence of passenger pick-up locations or a sequence of parking positions to wait for the next ride for the drivers. The objective of the mobile recommender system was to maximize the company level benefits. To meet this objective, the authors have proposed a Potential Travel Distance (PTD) function for evaluating each possible series of recommendations that is referred as a sequence. The proposed PTD function is monotonic which can be used to effectively prune the search space. Two algorithms namely the LCP and the SkyRoute algorithms were developed using the PTD function, for finding the recommended routes. The experimental results show that the proposed system can provide beneficial sequence of recommendations and the information that is extracted from the GPS traces can be used as a reference by the taxi drivers. The proposed recommendations when adhered to can effectively lead to an efficient use of energy.

Authors in [33] proposed a cruising system which they have termed as *pCruise*. The purpose of this system is to find efficient routes for the taxi drivers to ensure served demands are high whereas the idle cruising miles are minimized. Using the historical ride pick-up data by the taxis in the nearby spatial and temporal space, i.e., on the roads nearby the location the driver is waiting over a defined horizon, specific weights are given to each link. This is used in-order to recommend a path that maximizes the driver's chance of getting his next ride. The authors use a graph based approach to recommend the route which is actually a sequence of nodes and edges on the graph. They evaluated *pCruise* using GPS data from Shenzhen with a fleet size of 14000 taxicabs. The evaluation results showed that *pCruise* can result in a 42% reduction in the cruising miles of the taxis.

Dynamic Pricing

Dynamic pricing is a strategy to alter the supply-demand distribution by controlling the ride fare in a dynamic manner. Various pricing strategies have been used to mitigate the supply-demand imbalance problem. The central idea behind the dynamic pricing is that as the demand for the taxis increase, the taxi-fare can be increased. Such a strategy is called surge pricing. Due to surge in the taxi fares, it is expected that drivers from neighbouring regions will move towards the regions of higher demand. This will reduce the waiting time for the passengers and drivers will get a ride which will indirectly benefit the company.

Figure 1.8: Dynamic Pricing

The authors in [31] presented a dynamic general equilibrium model that was estimated using the NYC taxi trip data, for the taxi market. The authors state that the unique features of the taxi market are two fold. First, is what the authors term as the matching friction between the traders. This originates from the fact that the riders and the drivers (who are called the traders) cannot freely trade because of the constraints such as the distance between the rider and the passenger. If the distance is large, the friction between the respective rider and passenger is large as well. Secondly, there are regulatory limitations to enter the taxi market, i.e., a firm cap is at place on the total number of drivers that can enter the taxi market. The authors test the importance of these features by playing with the total fleet size, using different matching technologies and using different demand markets. This function exhibits increasing returns to scale, which is important to understand the impact of changes in this market and has welfare implications.

The authors in [30] reported an investigation in a regulated taxi market. A mechanism was proposed by which the demand and supply were matched in a regulated taxi market. In the proposed model, the passenger demand was assumed to be a function of factors such as the fare price as well as the taxi availability. The earnings of the taxi company was analyzed using the taxi utilization rate. The number of taxi vehicles supplied was assumed to be a function of the profits that the drivers can earn as well as what authors describe, the industrial structure. Taxi availability is quantified by the expected passenger waiting time, and taxi-utilization. Due to the complexity of the actual taxi market, especially its spatial and temporal aspects, the authors have made considerable amount of assumptions to make the analysis feasible. This study showed that the supply and demand are coupled, e.g., a higher level of taxi supply leads to a significantly drop in the passenger waiting time, hence improving the quality of service which attract more demand. Due to the higher demand value, the driver profits increase thus resulting in more supply. This supply-demand interaction can be explosive but eventually must damp out.

This paper [29] studied the impact of dynamic pricing on the supply of the labour. In this paper, the authors studied how the taxi drivers in the Uber platform react to the surge pricing. In order to maximize their profits, the drivers drive during the times when the markets are high and also flexibly adjust their schedules during the surge time. This study showed that surge pricing is effective in causing a supply surge in the area of high demand. They discussed the implications of this study over the driver profits, flexible work hours, and the

efficiency of dynamically-priced labor markets.

The authors in [7] state that if a taxi dispatch system is burdened due to excessive demands, this resulted in a situation where the taxis are very thinly distributed over the operating area. This forces the dispatcher to match the drivers and passengers that are far away from each other. The phenomena that arises here in which the taxis have to travel long distances is termed as the wild goose chase (WGC) phenomena. This leads to a wastage of the drivers' time and reduces their earnings. This leads to a critical situation in which the drivers start dropping out of the taxi market because of the excessive distances that they have to travel which also effects the quality of service in terms of low waiting times and lost demands. Authors propose that surge pricing is a solution to the wild goose chase problem because using higher prices provides an incentive for the drivers to stay in the market.

Authors in [15] proposed a methodology of supply-demand balancing by following the two strategies. First, they provided the information of the driver locations to the other drivers and secondly, they provided incentive to the drivers to go to the recommended locations. The objective was to optimize the set of waiting locations for the drivers. It was shown that implementing both the strategies was a NP Hard problem, so a near optimal solutions were found. After implementing the strategies to the real world data, it was observed that the performance of the proposed control methods showed a 20-80% increase in the response was observed in the drivers.

Authors in [16] proposed a methodology to jointly perform the vehicle rebalancing and optimal pricing strategy. At a higher level, the objective of the proposed model is to tackle the company profit maximization problem while ensuring the supply-demand balance in the operating area. The authors have used a fluid model to show that a stable equilibrium exists using the pricing policies. Further a joint optimization model was developed that uses the vehicle rebalancing and dynamic pricing as the supply-demand control strategy. Each approach was implemented separately and jointly in a sequential as well as simultaneous implementation manner. The simulation was performed using the New York Taxi dataset to access the performance of the proposed strategies.

1.1.3 Motivations and Thesis Organization

The motivation of this work is to build a new framework for an advanced taxi fleet management system. Most of the existing models focus on the company level benefits such as the served demands, company profits, quality of service, passenger waiting times etc. Some of the recent works have adopted a more driver-centric approach where they focus on increasing the driver booking rates by either recommender systems or vehicle rebalancing systems. However, the major challenge that has yet not been addressed completely is to integrate these multiple objectives into a single framework, i.e., most of the works either focus on improving the profits for the drivers by individually giving recommendations to them, or improve the fleet level benefits by using the operational policies such as vehicle rebalancing. However, the existing works don't consider the impact of the fleet coordination algorithms on the individual driver's profits, i.e., the existing frameworks do not incorporate the service fairness capabilities into their models. Mainly due to that reason, most of the proposed models have been developed for the fleet of autonomous taxis. This makes it difficult for the existing models to be extended to the fleet of vehicles with human drivers who shall only follow the instructions of the system if they see clear benefits in terms of higher profits.

In the proposed Coordinated Mobility-on-Demand (CMoD) system model we have incorporated various objectives within a single framework so that we can find an optimal balance between multiple fleet management objectives such as the company level benefits as well as the driver level benefits. The key features of the proposed model are, supply scheduling, i.e., serving the current taxi demands, vehicle rebalancing, i.e., redistributing the idle standing vehicles with an objective of maximizing the demands that can be served in the future and ensuring service fairness, i.e., improving the driver level benefits. We formulate the fleet management problem as a multi-objective constrained optimization problem incorporating the above mentioned features as the fleet management objectives, and then use a receding horizon based control to reduce the computational complexity while solving the optimization problem in real-time.

The Thesis is organized into five chapters, Chapter 2, 3, and 4 are the technical chapters that include the details of our model and chapter 5 presents the summary and the future works respectively.

In Chapter 2, we present the proposed Coordinated Mobility-on-Demand (CMoD) system model. First, we explain all the components of the model in detail. Second, we present the key elements pertaining to the proposed CMoD model, i.e., the supply scheduling, vehicle rebalancing and fleet control.

In Chapter 3, we identify the model parameters that are critical for the successful real-time implementation of the vehicle rebalancing model and using Upper Confidence Bound algorithm, we identify the optimal model parameters for the real-time implementation of vehicle rebalancing taking the passenger market into account.

Fleet coordination and operational strategies are more straight forward in implementation with regards to the autonomous mobility on demand system. However, implementing similar strategies in the system with human drivers is more challenging. The reason for this is that every driver is more concerned regarding his own profits. Hence, the challenge to implement fleet coordination strategies is to ensure that the drivers abide by the recommendations given by the central recommendation system. This is possible only if the drivers see a clear improvement in their profits.

In chapter 4, we present the extension of the proposed CMoD model to ensure that the driver profits are increased while serving the recommendations of the system. We formulate a multi-objective taxi fleet management problem, aiming for an optimal balance between the company profit and service fairness among individual drivers thus integrating all the key elements, i.e., supply scheduling, vehicle rebalancing, service fairness and fleet control in a single framework. We analyze the proposed strategy in a simulated taxi network derived from real taxi data from the database of New York City, and illustrate its effectiveness in comparison to other common practices and present the results of the tests and experiments that we had performed. We present three case studies to show the benefits of each objective term in the multi-objective fleet management model thus demonstrating the capability of the proposed model to ensure higher served demands as well as ensure service fairness among the drivers.

In the Chapter 5 we present the conclusions and future works.

Chapter 2

Coordinated Mobility-on-Demand Model

2.1 Introduction to CMoD Fleet Management

Contemporary mobility-on-demand systems that are partially coordinated are prone to under-utilization of the supply capacity. The primary reason for this inefficiency is that the taxi fleet is not owned by the ride-matching company, so the taxi drivers who seek to maximize their own profits and work according to their own schedule influence the supply distribution over the operating area. This leads to a supply-demand imbalance both spatially and temporally. This means the demand exceeds the taxi supplies in some regions, whereas the taxis are simultaneously standing idle in the other regions. With the implementation of the fleet coordination, it would be possible to have better control over the supply distribution using a centralized ride-matching and vehicle rebalancing system that seeks to maximize the company-level benefits while ensuring the service quality.

In this chapter we have presented the proposed coordinated mobility-on-demand system model, that improves the efficiency of taxi system operation in terms of higher capacity utilization by serving higher volume of demands. We provide a framework to anticipate and mitigate the future supply and demand imbalances using vehicle rebalancing policy. In this chapter, we also emphasize on the proposed improvements to the existing vehicle rebalancing models.

The chapter is organized as follows. First, we introduce the key challenges that are inherent to the coordinated mobility-on-demand systems and describe the hierarchical architecture of the proposed CMoD system. Then we describe the road network and the real-world taxi trips data that will form the base for the simulations. Next, we present a detailed formulation of the proposed mathematical model and discuss the key components of the model.

2.1.1 Problem Statement

The primary goal of the CMoD fleet management is to solve three problems, namely, the supply-scheduling, vehicle rebalancing and CMoD fleet control in each discrete time step.

- **Supply Scheduling:** In this problem, we calculate the maximum demand that can be served in each region using the available taxis. The taxis are dispatched from one region to the other in an occupied condition, i.e., with passenger on-board. This process is defined as the supply scheduling. Our objective is to maximize the number of taxis that can be scheduled depending on the current supply and demand distribution.
- **Vehicle Rebalancing:** In this problem, we decide how many taxis should we dispatch from one zone to the other, i.e., how should the vehicles be re-distributed with the objective of maximizing the demand that can be served in the near future while ensuring that the idle distance that the taxis have to travel is minimized.
- **CMoD Fleet control:** In order to implement the supply-scheduling and vehicle rebalancing problem in a sequential manner. We need a mechanism such as the Receding Horizon Control (RHC) to control the state of the system. We use RHC to update the number of vehicles that will be available in each region in the future time slots, as well as update the driver and system performance such as driver profits, net served demands etc.

Figure 2.1: Elements of CMoD Model

2.1.2 Hierarchy of CMoD System

Two-level hierarchy as shown in Fig 2.2 is proposed for the decision making in the Coordinated Mobility on Demand (CMoD) system. Operational policies such as the supply-demand balancing using vehicle rebalancing are implemented and controlled at the top level, whereas the bottom level consists of the driver allocation, i.e., selecting the best possible available driver for a pick-up request.

Figure 2.2: CMoD Hierarchy

The two levels of the hierarchy are explained as follows:

- **Top Level:** Operational policies like the supply scheduling and vehicle rebalancing decisions are made from a higher level of abstraction, i.e., at the regional level. This ensures reduction in the computational complexity. More specifically, we are concerned with the overall distribution of the supply and demand in each region. So, the decision that is involved at the top level is that how many taxis should be scheduled or recommended to go from one region to the other rather than which specific taxi should be dispatched.
- **Bottom Level:** Within each region, the driver allocation has to be done to the passenger pickup request. So, it is here that we need to decide that which specific taxi has to be scheduled depending on the distance of the driver from the passenger. Also, when we incorporate the service fairness into the system (as will be explained in the chapter 4), where we need to give the preference in scheduling to the drivers that are lagging in the profits or to those that have recently served a recommendation. These decisions are taken at the lower level giving separate importance to each driver.

In this chapter we shall look at the CMoD System's operation at the top level and we will dive into the driver specific aspects in Chapter 4.

2.1.3 Taxi Operational Network

We developed a simulated CMoD network based on the real data [8] from Manhattan. The network setup is shown below in Fig 2.3 where all the regions (taxi zones) with their centers are shown. The set of all the zones is denoted by Z such that $|Z| = n$ is the total number of zones in the city. There are total 66 zones which will be called the regions alternatively. Each taxi zone or region has the passenger pick up demands and the number of idle standing taxis, i.e., the supply associated with it.

The taxi operational network can also be viewed as a graph where each node of the graph represents a taxi zone or region and the neighbouring taxi zones are connected with an edge.

Figure 2.3: Manhattan Taxi Zones

2.1.4 Taxi Trips Data

The data that we have used for extracting the realistic taxi trips information (yellow cabs NYC) is as follows:

trip_id	day	start_time	PULocationID	end_time	DOLocationID	trip_time	trip_distance	trip_fare
0	0	0	68	13	170	13	2.2	8
1	0	0	79	13	162	13	3	10
2	0	0	75	3	75	3	0.6	4
3	0	0	236	5	262	5	0.89	4.725
4	0	0	237	4	141	4	0.81	4.525
5	0	0	161	7	142	7	1.58	6.45
6	0	0	48	16	263	16	3.49	11.225
7	0	0	161	19	263	19	4	12.5
8	0	0	161	19	263	19	4	12.5
9	0	0	263	11	151	11	2.6	9
10	0	0	107	39	244	39	9.9	27.25

Figure 2.4: Yellow Taxi Trip Data

- *trip_id*: A unique ID that is used for representing a trip.
- *day*: The day of the month $\in [0,30]$.
- *start_time*: Time slot in which the trip started.
- *PULocationID*: The ID of the taxi zone in which the trip began.
- *end_time*: Time slot in which the trip ended.
- *DOLocationID*: The ID of the taxi zone in which the trip ended.
- *trip_time*: Number of time slots it took to complete the trip.
- *trip_distance*: The distance travelled during the trip in km.
- *trip_fare*: The trip fare is the metered fare in \$.

Data Processing

The above shown table was extracted from the the raw taxi trip data [8]. We had to perform the following transformations on the raw data to get the table as seen in Fig 2.4

- Time slots: Trip start and end time in the raw data is in form of a timestamp. For our formulation, we chose a time slot duration of 1 min. The actual time was then mapped to the time slot, e.g., if the trip started any time between 00:00:00 to 00:00:59, we mapped it to time slot 0. The width of the time slot can be decided according to the requirement.
- Trip Fare: We used a standard meter fare as explained in [8] to evaluate the trip fare. The formula to calculate the basic fare is per km cost x trip distance + fixed cost.

2.1.5 Components of CMoD System

The key components that form the integral part of the CMoD system are as follows:

Supply Matrix ($n \times K$)

Supply matrix $V(k)$ stores the number of taxis that are standing idle, i.e., $v_j(k) \in V(k)$ is the number of taxis standing idle in zone $j \in Z$ in slot $k \in [1, K]$. Supply matrix is updated at the end of every time slot k as a part of the fleet control operation.

As shown in the figure below, let the current time slot be k . Let there be three regions and 6 taxis in total. The figure below shows that there are 2 taxis in region 1, 1 taxi in region 2 and 2 taxis in region 3 during the current time slot. There is one taxi that is en-route to region 2 and will reach there during time slot $k+3$. The figure below gives a snapshot of the values of the supply matrix.

Figure 2.5: Supply Matrix

Demand Matrix ($n \times n$)

We use a matrix representation to store the passenger demands. The demand matrix, which is denoted by $A(k)$, stores the number of requests that appeared during time slot k , where $a_{ij}(k) \in A(k)$ is equal to the number of requests during the time slot k to go from zone $i \in Z$ to $j \in Z$ which is denoted by $i \rightarrow j$. Here, i is the zone from which the trip started and j is the zone in which the trip ended.

As seen in the matrix representation below, we store the inter-region travel demand in form of a matrix. Where, $a_{ij}(k)$ implies the number of requests to travel from region i to region j .

$$A(k) = \begin{bmatrix} a_{11}(k) & a_{12}(k) & a_{13}(k) & a_{14}(k) & a_{15}(k) & \dots & a_{1n}(k) \\ a_{21}(k) & a_{22}(k) & a_{23}(k) & a_{24}(k) & a_{25}(k) & \dots & a_{2n}(k) \\ a_{31}(k) & a_{32}(k) & a_{33}(k) & a_{34}(k) & a_{35}(k) & \dots & a_{3n}(k) \\ a_{41}(k) & a_{42}(k) & a_{43}(k) & a_{44}(k) & a_{45}(k) & \dots & a_{4n}(k) \\ & & & \vdots & & \dots & \\ a_{n1}(k) & a_{n2}(k) & a_{n3}(k) & a_{n4}(k) & a_{n5}(k) & \dots & a_{nn}(k) \end{bmatrix}$$

Algorithm to extract the Demand Matrix

Note that, n ($= 66$) is the total number of taxi zones in Manhattan and K ($= 1440$ because the time slot duration is 1 min so 24×60 timeslots) is the total number of time slots. R is the index of the last row of the trip data table. Each zone is given a zone id $\in [1,66]$, i.e., the actual PULocationID and DOLocationID (as given in data table) was re-defined to a new unique value in $[1,66]$.

Algorithm 1

```

1: procedure LOAD TRIP DATA ▷ data is shown in Fig 2.3
2:
3:    $A \leftarrow \text{zeros}[n, n, K]$  ▷  $A \leftarrow$  three dimensional matrix of zeros
4:
5:   while  $r \leq R$  do
6:      $i \leftarrow$  index of zone in which the trip started for row  $r$ 
7:      $j \leftarrow$  index of zone in which the trip ended for row  $r$ 
8:      $k' \leftarrow$  time slot in which the trip ended
9:
10:     $A[i, j, k'] := A[i, j, k'] + 1$ 
11:  return  $A$  ▷ Return the updated demand matrix

```

Anticipated Demand Vector ($n \times 1$)

In-order to implement the vehicle rebalancing we need to anticipate the passenger demands, i.e., the number of ride requests that will appear in each zone $i \in Z$ during a finite prediction horizon ω . This anticipated demand will be used to plan the vehicle rebalancing strategy, i.e., to re-distribute the idle standing vehicles (if required) so that more demands can be served in the near future.

Passenger demand prediction is generally modelled as a time-series forecasting problem. Recurrent Neural Networks or ARIMA models are used to predict the passenger pick-up demands. In an ideal scenario, i.e., with a perfect predictor the predicted demand in zone $i \in Z$ generated in time slot k over a prediction horizon ω is same as the ground truth from the data [8] given by:

$$\mathcal{D}_i(k, \omega) = \sum_{t=k+1}^{k+\omega} \sum_{j=1}^n a_{ij}(t) \quad (2.1)$$

All the predictors are subject to prediction inaccuracies due to the following two main reasons:

- The passenger demand prediction may not always show a perfectly repeatable pattern. Hence, many authors model the passenger request arrival as a stochastic process. Particle filters and time varying Poisson processes have been used to generate the passenger arrivals. Hence, it's only possible to get the expected value of the passenger demand even if the underlying stochastic process is completely known.
- Even if the passenger demand pattern is perfectly repeatable, there is another factor of request drop-out that is generally difficult to model. The reason being that the passenger may cancel the request if they think the price is too high or they have to wait for longer period of time.

We want to assess impact of the demand uncertainties on the performance of the CMoD model. In order to achieve that we use the actual demand signal that can be extracted using the equation 2.1 as a baseline. Then we add a Gaussian noise to that signal in order to incorporate these uncertainties. This mechanism is explained in more detail in the next subsection. We use the noisy prediction for a given value of noise parameter α , denoted by $\widetilde{\mathcal{D}}_j(k, \omega, \alpha)$ as predicted future demand, while solving the vehicle rebalancing problem.

Uncertainty: Noise Parameter (α)

Demand forecasting is prone to uncertainties that increase with the prediction horizon window ω , i.e., the prediction accuracy falls as the prediction horizon increases. In-order to replicate this, we introduce a zero mean Gaussian noise whose standard deviation is a monotonically increasing function of the prediction horizon. Demand Forecast is given as follows:

$$\widetilde{\mathcal{D}}_j(k, \omega, \alpha) = \sum_{t=k+1}^{k+\omega} \left(\sum_i a_{ji}(t) + N(0, g(t)) \right) \quad (2.2)$$

Where,

- $a_{ji}(t)$: actual demand to go from region $j \rightarrow i$
- $N(0, g(t))$: gaussian noise

We used a linear function for standard deviation, $g(t) = \alpha t$. This ensures that the deviation of the predicted signal from the baseline or the actual signal increases with the increase in the prediction horizon. Fig 2.6 shows the effect of α on the demand prediction across Manhattan zones in a time duration of 10 mins. The signal shown in blue is the actual number of passenger requests in taxi zone [1-66] for a duration of 10 mins. Using the noise parameter α , we can control the deviation of the predicted signal from the actual signal.

Figure 2.6: Noise Analysis

Anticipated Supply Vector ($n \times 1$)

For vehicle rebalancing, we need to plan if there is a requirement to dispatch taxis from one region to the other by predicting the upcoming passenger pick-up demand in each region. For an effective vehicle rebalancing, i.e., rebalancing the vehicles at a minimum cost, we need to know the number of vehicles that are en-route to every zone $i \in Z$ and are expected to reach during time $k+1$ to $k+\omega$. This is necessary because we don't need to dispatch the taxis to a region if there are sufficient number of taxis en-route to that region. The estimate of the total number of vehicles that are en-route to any region is defined as the anticipated supply. In our model, we record the number of vehicles in each region during any particular time-slot of the day in the Supply Matrix $v_i(t) \in V$. So the anticipated supply in region i during time duration $k+1$ to $k+\omega$ is:

$$s_i(k, \omega) = \sum_{t=k+1}^{k+\omega} v_i(t) \quad (2.3)$$

Example

The following simplified example gives an idea about the significance of anticipating the supply and demand correctly. In this example, we have three regions and the current time slot is k . Starting from this time slot on-wards, if we look ahead with a prediction horizon ω , we anticipate 1 passenger pickup request in region 1 and region 3. We also anticipate that there shall be no upcoming demands in region 2. There is one idle standing taxi in region 2 and 1 taxi is en route to region 3. We anticipate that the taxi that is en route to region 3 will reach there during the time interval $k+1 \rightarrow k+\omega$. So the anticipated supply in region 3 is 1, i.e., we anticipate there will be 1 taxi in region 3, which is currently en route. Now, we have two choices for the taxi that is standing idle in region 2. Either, we can dispatch that taxi to region 1 or to region 3. However, dispatching the idle standing taxi to region 1 may lead to higher served demands in the future. So, we dispatch the idle taxi to region 1.

Figure 2.7: Anticipated Supply and Demand

Distance Matrix ($n \times n$)

The trip distance from region $i \in Z$ to region $j \in Z$ may vary depending on the actual GPS coordinates of the location where the trip started and where the trip ended. We calculate the average taxi trip distance from region $i \in Z$ to region $j \in Z$ using the trip records given in the data table. Distance matrix D records the average trip distance from region i to j which is denoted by $d_{ij} \in D$. We use the average trip distance in order to estimate the cost of the fuel that will be incurred on average while travelling from region i to j .

Travel Time Matrix ($n \times n$)

Similar to the distance matrix, we use the travel time matrix T that records the average travel time along the trip from region i to j which is denoted by $t_{ij} \in T$. The average travel time is used to estimate the time it will take to travel from region i to j . The travel time estimation is essential because while planning the vehicle rebalancing, we need to ensure that the taxis that are dispatched from one region to the other should be able to reach there within the prediction horizon.

Reachability Matrix ($n \times n$)

Reachability matrix R is closely related to the travel time matrix T and a prediction horizon ω . Such that $\rho_{ij} \in R$ is given by:

$$\begin{aligned} \rho_{ij} &= 1 & \text{if } t_{ij} \leq \omega \\ \rho_{ij} &= 0 & \text{if } t_{ij} > \omega \end{aligned}$$

Reachability index $\phi = \frac{1}{n^2} \sum_{i,j|i \neq j} \rho_{ij}$ is the fraction of inter-zone travel that is possible within prediction horizon ω

Algorithm 2

```

1: procedure LOAD TRIP DATA ▷ data is shown in Fig 2.3
2:    $D, T, N \leftarrow \text{zeros}[n, n]$  ▷ D, T ← two dimensional matrix of zeros
3:   while  $r \leq R$  do
4:      $i \leftarrow$  index of zone in which the trip started for row  $r$ 
5:      $j \leftarrow$  index of zone in which the trip ended for row  $r$ 
6:      $N[i, j] := N[i, j] + 1$  ▷  $N[i, j]$  is # of trips from region  $i$  to  $j$ 
7:
8:      $D[i, j] := \frac{(N[i, j] - 1) \cdot D[i, j] + d[r]}{N[i, j]}$  ▷  $d[r]$  is the distance travelled for trip  $r$ 
9:      $T[i, j] := \frac{(N[i, j] - 1) \cdot T[i, j] + t[r]}{N[i, j]}$  ▷  $t[r]$  is the travel time for trip  $r$ 
10:  return  $D, T, N$  ▷ Return the updated matrices

```

Profit Matrix ($n \times n$)

Trip profit r_{ij} from a taxi ride is based on a standard meter fare. Fig 2.8 shows a snapshot from the website of Taxi and Limousine Commission [8] that shows how the standard meter fare for the yellow taxi is calculated.

For our experiments we only consider two components, i.e., the per mile charge (μ) and initial charge (δ) both in dollar. Detailed expression for the trip profit is explained in subsection 2.2.3.

Figure 2.8: Source: NYC Taxi and Limousine Commission [8]

Cost Matrix ($n \times n$)

For trip cost c_{ij} calculation we need mileage χ (gallons/mile) of taxis and fuel cost ξ (dollars/gallon). Trip cost is proportional ($c_{ij} \propto d_{ij}$) to the trip-travel distance d_{ij} . Detailed expression for the trip cost is explained in subsection 2.2.3. We get the mileage information from the specs of a popular yellow taxi model [21]. Note that we have merely used this taxi as a reference and for the purpose of uniformity. A taxi fleet has different types of taxis and it is impractical to use the different mileage information for each taxi. We get the fuel cost information from the US Energy Information Administration [22].

2.2 Mathematical Model for CMoD Fleet Management

CMoD fleet management is modelled as a sequential decision making process. Discrete time steps are used to test the model. The overall problem can be decomposed into three sub-problems which are solved at each time step k . The three sub-problems are as follows:

- *Supply Scheduling*: Supply scheduling implies the total number of taxis that can be dispatched from one region to the other in an occupied condition, i.e., with a passenger on board. The objective of solving the supply scheduling problem is to maximize the number of taxis that can be scheduled which depends on the number of idle standing taxis as well as the passenger pick-up requests in each region.
- *Vehicle Rebalancing*: The objective of this problem is to maximize the demands that can be served in the near future, i.e., within a finite prediction horizon by re-distributing the idle standing vehicles. We use the anticipated demand and supply to perform the vehicle rebalancing.
- *CMoD Fleet Control*: We need to update the state of the system such as the position of the taxis, number of served requests, driver profits, etc at the end of each time slot using the Receding Horizon Control (RHC) framework. This, is termed as the CMoD Fleet Control.

Figure 2.9: Coordinated Mobility-on-Demand Objectives

2.2.1 Supply Scheduling

During each time slot k , we need to calculate the maximum demand that can be served with the available taxis. Supply scheduling refers to dispatching the idle standing taxis from one region to the other in an occupied condition, i.e., with the passenger on board. We are given the number of idle standing taxis and the current passenger pick-up requests in each region. Our objective is to build a mathematical model that can determine the following:

- The maximum number of taxis that can be scheduled in each region.
- The dispatch solution, i.e., how many taxis can be dispatched from one region to the other.
- The updated distribution of the taxis in each region, i.e., how many taxis will be standing idle in each region after implementing the dispatch solution.

The figure shown below gives a basic idea of the supply scheduling process. As seen in the figure, there were two taxis standing idle in region 2, two taxis standing idle in region 1 and there is no idle taxi in region 3. There are two passenger pick-up requests in the region 2, one request in region 1 and 4 requests in region 3. It can be seen visually that the total number of requests that can be served in this case are 3, i.e. one request in region 1 and 2 requests in region 2. The request in region 1 has destination in region 3. One of the request in region 2 has destination in region 1 and the other request has destination in region 3.

Figure 2.10: Supply Scheduling

Problem Statement

Given the information of the supply, i.e. number of taxis in each region and the demand, i.e., the number of passenger pickup requests in each region, we need to determine the maximum number of taxis that can be scheduled, i.e., dispatched from one region to the other in an occupied condition.

The number of passenger requests are recorded in the demand matrix $A(k)$ where, $a_{ij}(k) \in A(k)$ are the number of requests to go from region $i \in Z$ to $j \in Z$.

The total number of taxis that are standing idle in region $i \in Z$ can be extracted from the supply matrix $V(k)$ and is given by $v_i(k) \in V(k)$.

Corresponding to each trip from region i to j , there is a profit associated with that travel. This profit value is stored in the profit matrix and is denoted by r_{ij}

Decision Variable

Decision variable $x_{ij}(k)$ determines the number of taxis that can be scheduled to go from region i to j . The variable $x_{ij}(k) \in R_+$, i.e., can take any positive real value.

Constraints

1. Total number of taxis that can be scheduled to go from region i to all the other regions, is no more than the number of taxis standing idle in region i . The total number of taxis standing idle in region i during time slot k is equal to $v_i(k)$. The total number of taxis that are scheduled to go from region i to the other regions are $\sum_{j=1}^n x_{ij}(k)$. So, the constraint that ensures that the taxis scheduled from one region to the other are less than the number of taxis standing idle in that region is given by the equation:

$$\sum_{j=1}^n x_{ij}(k) \leq v_i(k) \quad (2.4)$$

2. Total number of taxis dispatched from region $i \in Z$ to region $j \in Z$ is no more than the number of passenger requests to go from region $i \in Z$ to $j \in Z$. The number of passenger requests to go from region i to j such that the pick-up request appears at the time slot k is equal to $a_{ij}(k)$. So, the constraint that ensures that the taxis scheduled from one region to the other are less than the number of passenger requests to go from region i to j is given by the following equation:

$$x_{ij}(k) \leq a_{ij}(k) \quad (2.5)$$

Problem Formulation

We formulate the supply scheduling problem as a linear program where we try to maximize the total profit that can be earned by scheduling the idle standing taxis. We solve the supply scheduling problem during each time slot k to determine the value of decision variable $x_{ij}(k) \forall i, j$.

The information that is available with us is the number of taxis that are standing idle in each region, i.e., $v_i(k)$, the number of passenger requests to go from region i to j , during time slot k , which is given by $a_{ij}(k)$ and the profits that can be earned through a trip from region i to j , which is given by r_{ij} .

The problem formulation, for solving the supply scheduling problem is as follows:

$$\begin{aligned}
 \max_x \quad & \sum_{i,j=1}^n r_{ij}x_{ij}(k) \\
 \text{s.t.} \quad & \sum_{j=1}^n x_{ij}(k) \leq v_i(k) \quad \forall i \in Z \\
 & x_{ij}(k) \leq a_{ij}(k) \quad \forall i, j \in Z
 \end{aligned} \tag{2.6}$$

- $x_{ij}(k)$ is a decision variable that decides the number of taxis that should be dispatched from i to j and r_{ij} is the profit that will be earned while travelling from i to j .
- So, $r_{ij} \cdot x_{ij}$ is the profit that can be earned after dispatching x_{ij} number of taxis from region i to region j .
- The expression $\sum_{i,j=1}^n r_{ij}x_{ij}(k)$ implies the total profit that will be earned by scheduling the available taxis across all the regions.
- Objective: We need to maximize the total profit that can be earned by scheduling the taxis to go from one region to the other.
- Supply Constraint: In every region $i \in Z$, the total number of taxis that get scheduled are less than the total number of taxis that are standing idle in that region, which is denoted by $v_i(k)$.
- Demand Constraint: In every region $i \in Z$, the total number of taxis that get scheduled to go from region i to j are less than the number of passenger requests to go from region i to j , which is denoted by $a_{ij}(k)$.

2.2.2 Vehicle Rebalancing

During each time slot k , after finishing the supply scheduling. We are left with some idle standing taxis in each region that didn't get scheduled during the supply scheduling process. We also have some taxis that are en-route to drop the passenger who is on board. Using the information of the idle standing taxis and the taxis that are en-route a region, we can anticipate the supply in each region. This anticipated supply is given by the following equation:

$$s_i(k, \omega) = \sum_{t=k+1}^{k+\omega} v_i(t)$$

Note that the term $v_i(t)$ includes the updated value of the number of taxis in each region $i \in Z$ after the supply scheduling.

We also need to predict the number of requests in each region within the prediction horizon. Since, we already have the ground truth ,i.e., the number of requests that appeared in each region. We introduce a noise to these predictions in order to replicate a demand predictor. The noise level in the demand prediction is controlled using a noise parameter α as explained in the previous section. The predicted demand is given by the following equation:

$$\widetilde{\mathcal{D}}_j(k, \omega, \alpha) = \sum_{t=k+1}^{k+\omega} \left(\sum_i a_{ji}(t) + N(0, g(t)) \right)$$

Where, $g(t) = \alpha t$, i.e., the noise value increases linearly with time. The objectives of the vehicle rebalancing model are as follows:

- Anticipate the passenger pick-up demand, i.e., the number of requests that will appear in each region during the prediction horizon and the taxi supply, i.e., the number of taxis that will be available in each region $j \in Z$ within the prediction horizon.
- Using the anticipated demand and supply, generate a dispatch solution, i.e., how many idle standing taxis should be dispatched from one region to the other in-order to maximize the demands that can be served in the near future, i.e., within the prediction horizon.

Example

The figure shown below gives a basic idea of the vehicle rebalancing process. For rebalancing the vehicles among the regions, we need to anticipate the total number of passenger requests that will appear in each region and total number of taxis that will be available in each region during the prediction horizon period ,i.e., k to $k + \omega$. As seen in the figure, lets assume that the anticipated supplies are two vehicles in region 1, i.e., one vehicle which is en-route to region 1 and one vehicle which is standing idle in region 1 as shown in the following figure. For region 2 the anticipated supply is two, i.e., two vehicles that are standing idle in region 2. For region 3, it is anticipated that there shall be no available taxi in the prediction horizon. The predicted demand, i.e., the number of passenger pick-up requests in region 1 is one, region 2 is two and region 3 is 4. It can be seen visually that if we dispatch one idle standing taxi from region 1 to region 3, we might be able to serve higher demands in the near future as compared to the situation where we don't dispatch any vehicle from one region to the other.

Figure 2.11: Vehicle Rebalancing

Problem Statement

Given the information of the anticipated demand and supply in each region $j \in Z$, we need to determine the number of taxis that should be sent from one region to the other in-order to balance the supply and demand gap in each region in the near future. Our objective is to maximize the expected demand that can be served in the near future.

The total number of taxis that are standing idle in region $i \in Z$ after supply scheduling can be extracted from the updated supply matrix $V(k)$ and is given by $v_i(k) \in V(k)$.

The travel time for the trip from region i to j is recorded in the travel time matrix such that $t_{ij} \in T$ is the time taken to travel from region i to j .

The anticipated supply, i.e., total number of vehicles that are en-route a particular region is denoted by $s_j(k, \omega)$.

The anticipated demand, i.e., the number of passenger requests that are expected to arrive in region j within the prediction horizon ω is given by $\widetilde{\varphi}_j(k, \omega, \alpha)$

Decision Variables

- $y_{ij}(k)$ is the number of taxis that are recommended to go from region i to region j starting at the time slot k . For vehicle rebalancing, we need to first anticipate the demand in each region, i.e., predict the number of passenger requests that are expected to appear in each region in the near future. We also need to anticipate the supply, i.e., the number of taxis that are en-route region $j \in Z$. Based on the anticipated demand and supply in each region, we need to decide how many taxis should be dispatched from one region to the other.
- $z_j(k)$ is the expected number of requests that can be served in region $j \in Z$ depending on the anticipated demand and supply in each region. The number of requests that can be served in the near future is no more than the actual number of passenger requests and idle standing taxis in each region $j \in Z$.

Constraints

1. The total number of taxis that can be recommended to go from region i to all the other regions, is no more than the number of taxis standing idle in region i after the supply scheduling has been completed. The updated total number of taxis standing idle in region i during time slot k is equal to $v_i(k)$. The total number of taxis that are scheduled to go from region i to other regions are $\sum_{j=1}^n y_{ij}(k)$. So, the constraint that ensures that the taxis scheduled from one region to the other are less than the number of taxis standing idle in that region is given by the equation:

$$\sum_{j=1}^n y_{ij}(k) \leq v_i(k) \quad (2.7)$$

2. We need to ensure that only the taxis that can reach from i to j within the prediction horizon ω should be dispatched. The time taken to travel from region i to j is given as t_{ij} . The prediction horizon is ω . The following equation can be used to satisfy this constraint:

$$(t_{ij} - \omega) \cdot y_{ij}(k) \leq 0 \quad (2.8)$$

3. The anticipated number of requests that can be served in region j during the time period $k+1$ to $k+\omega$ which is denoted as z_j , should be no more than the anticipated supply $s_j(k, \omega)$ plus the total number of vehicles recommended to go to the zone j . The following equation can be used to satisfy this constraint:

$$z_j(k) \leq s_j(k, \omega) + \sum_i y_{ij}(k) \quad (2.9)$$

4. The anticipated number of requests that can be served in region j during the time period $k+1$ to $k+\omega$ which is denoted as z_j , should be no more than the expected number of requests that will appear in the zone j . The expected number of requests that appear in zone j is influenced by the noise parameter α . The following equation can be used to satisfy this constraint:

$$z_j \leq \widetilde{\wp}_j(k, \omega, \alpha) \quad (2.10)$$

Problem Formulation

We formulate the vehicle rebalancing problem as a multi-objective constrained optimization problem. This problem involves anticipating the demand in each region over a finite time horizon in the future and dispatching the idle standing taxis after supply scheduling from one zone to the other if required, with the objective of maximizing the number demands that can be served in near future at a minimum cost. The mathematical formulation for this problem is as follows:

$$\begin{aligned}
 \max_{z,y} \quad & \sum_{j=1}^n z_j - \beta \sum_{i,j=1}^n c_{ij} y_{ij}(k) \\
 \text{s.t.} \quad & \sum_{j=1}^n y_{ij}(k) \leq v_i(k) \quad \forall i \in Z \\
 & (t_{ij} - \omega) \cdot y_{ij}(k) \leq 0 \quad \forall i, j \in Z \\
 & z_j \leq s_j(k, \omega) + \sum_i y_{ij}(k) \quad \forall j \in Z \\
 & z_j \leq \widetilde{\phi}_j(k, \omega, \alpha) \quad \forall j \in Z
 \end{aligned} \tag{2.11}$$

- The decision variable y_{ij} is the total number of taxis that have been recommended to go from region i to j and c_{ij} is the fuel cost to travel from region i to j .
- So, the term $c_{ij} \cdot y_{ij}$ implies the total fuel cost of travelling from region i to j .
- The total fuel cost that is incurred due to the inter zonal travel while serving the recommendations is given by the equation:

$$\sum_{i,j=1}^n c_{ij} \cdot y_{ij}(k)$$

- Decision variable $z_j(k)$ is the anticipated number of requests that will get served during the prediction horizon ω .
- Using the above shown objective function, we maximize the total expected number of served demands, i.e., $\sum_{j=1}^n z_j$ at a minimum fuel cost value, i.e., $\sum_{i,j=1}^n c_{ij} y_{ij}(k)$.

2.2.3 CMoD fleet control

We apply a receding horizon strategy when carrying out optimization in an iterative manner. The supply scheduling problem is solved during each time-slot. Whether to implement the vehicle rebalancing in time slot k is decided by the vehicle rebalancing policy. Solving either of the problems leads to a change in the system parameters, e.g., the spatial distribution of the idle standing vehicles. At the end of each time slot, we update the system parameters and record the performance indicators in the following manner:

Supply Distribution

After scheduling or rebalancing, the distribution of the taxis in each zone changes. The taxis are dispatched from zone $i \rightarrow j$ in an occupied or unoccupied condition. If the travel time from, $i \rightarrow j$ is given by t_{ij} , the aggregate number of idle vehicles in zone j during time slot k' which is denoted by $v_j(k')$ is update in the following manner:

$$v_j(k') := v_j(k) + \delta(k) \sum_{i=1}^n y_{ij}(k) + \sum_{i=1}^n x_{ij}(k) \quad (2.12)$$

where, $k' = k + t_{ij}$

Binary variable $\delta(k)$ is 1 if vehicle rebalancing is planned for time slot k ; and 0 otherwise. The value of $\delta(k)$ is decided by the rebalancing policy and the passenger demand market as explained in the next section.

Served Demand Ratio (SDR)

SDR is a performance indicator that accesses the company level performance of the CMoD fleet. The primary objective of a CMoD system is to serve as much demand as possible. We extract the actual demands, i.e., the passenger trip requests from $i \rightarrow j$, that appeared in time slot k from the real trip records [8]. We also record the served-demands during time slot k after solving the *Supply Scheduling* problem. We use the ratio of aggregated served demands to the actual demands as a performance measure of our model. This ratio is termed as the Served Demand Ratio (SDR) which is given by the following equation:

$$SDR = \sum_k \sum_{i,j} \frac{x_{ij}(k)}{a_{ij}(k)} \quad (2.13)$$

Profit and Cost

There is a monetary value involved while traveling from zone $i \rightarrow j$. In case the taxi is occupied, the value is positive because of ride fare and in case the taxi is empty the monetary value is negative because of fuel cost.

Trip income from a taxi ride is based on a standard meter fare. We only consider two components, i.e., the per mile charge (μ) and fixed charge ζ . Other components like tolls, tips, surcharges etc are not considered.

For trip cost calculation we need the mileage χ (gallons/mile) of taxis and fuel cost ξ (dollars/gallon). Money m_{ij} that is made by a driver from a taxi ride from i to j with trip distance d_{ij} is given by the following equations:

$$m_{ij} = \mu * d_{ij} + \zeta \quad (2.14)$$

However, while travelling on this trip, the driver has to bear some cost c_{ij} in terms of fuel that is used.

$$c_{ij} = d_{ij} * \chi * \xi \quad (2.15)$$

Thus the driver's profit from this trip ride is: r_{ij} where,

$$\begin{aligned} r_{ij} &= m_{ij} - c_{ij} \\ r_{ij} &= (\mu - \chi\xi) d_{ij} + \zeta \end{aligned} \quad (2.16)$$

As a performance measure, we calculate the total profit made by the fleet at the company level. The total profit p that is earned by the fleet is the summation of the profits for the scheduling at each time slot k .

$$p = \sum_k \sum_{i,j} r_{ij} x_{ij}(k) \quad (2.17)$$

Vehicle rebalancing has a negative impact in-terms of higher operational costs. The reason is that the idle standing vehicles are dispatched from one zone to the other in order to serve the higher demands in the near future. The total cost κ is given by:

$$\kappa = \sum_k \sum_{i,j} c_{ij} y_{ij}(k) \quad (2.18)$$

Recommendation Efficiency η

We record the number of recommendations that have been served by the system. The aggregate number of recommendations are given by the expression:

$$\sum_k \sum_{i,j} y_{ij}(k)$$

In order to access the efficiency of the recommender system, we need to calculate how *useful* were the recommendations given by the vehicle rebalancing model. The primary reasons for a recommendation not being useful are as follows:

- The demand predictors such as the time-series forecasting models are always prone to inaccuracies because of the random nature of the passenger request arrivals.
- The predicted demand is subject to drop-out because of the finite patience level of the passengers may cancel the request if not served within a finite time after the arrival.

In case, the number of demands that were served during the prediction horizon, i.e., $k+1 \rightarrow k+\omega$ exceeds the number of expected supplies without rebalancing, we consider such recommendations as *useful*.

The reason for this is that if there was no rebalancing, the number of requests that would have been served, would have been at most equal to the already idle standing vehicles plus the incoming vehicles. So, we calculate the excessive number of requests that were served by subtracting the anticipated supply from the actual served demand.

The ratio of the *useful* recommendations to the total recommendations that were made to go to region j at time slot k is given as follows:

$$\eta_j(k) = \frac{\sum_{t=k+1}^{k+\omega} \sum_i x_{ji}(t) - s_j(k, \omega)}{\sum_i y_{ij}(k)} \quad (2.19)$$

Recommendation efficiency η :

$$\eta = \frac{\sum_k \sum_j \left(\left(\sum_{t=k+1}^{k+\omega} \sum_i x_{ji}(t) \right) - s_j(k, \omega) \right)}{\sum_k \sum_{i,j} y_{ij}(k)} \quad (2.20)$$

2.2.4 Model Variable Summary

Model Variables	
Variable	Description
$x_{ij}(k)$	no. of taxis scheduled to go from $i \rightarrow j$ during k
$y_{ij}(k)$	no. of taxis recommended to go from $i \rightarrow j$ during k
$z_j(k)$	expected number of served requests during the prediction horizon $k \rightarrow k + \omega$
$\delta(k)$	binary decision variable that decides if rebalancing is planned for time slot k
$a_{ij}(k)$	no. of passenger requests from $i \rightarrow j$ during k
$v_j(k)$	no. of vehicles standing in region j during time slot k
r_{ij}	ride fare for travelling from $i \rightarrow j$
c_{ij}	cost for travelling from $i \rightarrow j$
t_{ij}	average trip time for travelling from $i \rightarrow j$
d_{ij}	average trip distance for travelling from $i \rightarrow j$
ρ_{ij}	binary variable that decides if it is possible to reach from zone $i \rightarrow j$ during $k \rightarrow k + \omega$
μ	per mile charge in dollars/mile
ξ	fuel cost in dollars/gallon
ζ	fixed charge in dollars
χ	mileage in gallons/mile
f	balancing frequency
ω	prediction horizon
η	recommendation frequency
κ	fraction of company's share
φ	reachability index
$s_j(k, \omega)$	predicted supply in region j during $k \rightarrow k + \omega$
$\widetilde{\varphi}_j(k, \omega, \alpha)$	predicted demand in region j during $k \rightarrow k + \omega$

Table 2.1: CMoD Model Variables

2.3 Analysis of the Impact of Vehicle Rebalancing

In order to visualize the impact of the proposed supply-demand balancing policy, we recorded the number of requests that were observed and served for the morning peak, i.e., from time 06:00:00 to 10:00:00.

Two models were compared to check the performance of the proposed CMoD model. These models are explained as follows:

- Model 1: In this model, we only performed supply scheduling by solving the supply scheduling problem 2.6. In this problem, we evaluate the number of taxis that need to be dispatched from one zone to the other in an occupied condition. However, once the taxis reach the destination, they wait in their respective regions till the time they are scheduled again.
- Model 2: In this model, we first perform the supply scheduling by solving the problem 2.6. Then, for the remaining taxis, we perform the vehicle rebalancing by solving the problem 2.11. So, after dispatching the taxis from one region to the other in an occupied condition (supply scheduling), the taxis are dispatched in an unoccupied condition (vehicle rebalancing) so that the demands that can be served in the future are maximized.

We recorded the total number of passenger requests that appeared in each region during the above mentioned time duration which is given by the expression $\sum_{t=k+1}^{k+\omega} \sum_{i=1}^n a_{ji}(t)$. We performed two experiments with the same taxi fleet size and passenger demands. In the first experiment, we used the Model 1 and recorded the number of requests that were served in each region which is given by $\sum_{t=k+1}^{k+\omega} \sum_{i=1}^n x_{ji}(t)$. In the second experiment, we used Model 2 with the same initial distribution of taxis over the regions, and same passenger pickup demand, and again recorded the number of requests that were served in each region. The gap between the actual and the served demand is the lost demand which is given by the following equation:

$$\sum_{t=k+1}^{k+\omega} \sum_{i=1}^n a_{ji}(t) - \sum_{t=k+1}^{k+\omega} \sum_{i=1}^n x_{ji}(t) \quad (2.21)$$

Fig 2.12 shows a heat map corresponding to the gap between the actual demand and the served demand. The regions in red have high value of the lost demand. Whereas, the regions in green have low value of the lost demand. The figure on the left shows the gap in the served demand and the actual demand for Model 1, i.e., without vehicle rebalancing and the figure on the right show the gap for Model 2, i.e., with vehicle rebalancing. It is evident from the plot that using the proposed vehicle balancing model mitigates the supply-demand imbalance using the same number of taxis.

Figure 2.12: Impact of Vehicle Rebalancing

2.4 Conclusion

In this chapter, we first introduced the key challenges pertaining to the CMoD system, namely the supply scheduling, vehicle rebalancing, and fleet control, and explained the taxi operational network as well as the real-world taxi trip data that we have used for the simulation purpose. Then we presented a detailed explanation of the proposed CMoD model. Firstly, we presented the key components of the model. Secondly, we elaborated the supply scheduling model which is a linear programming-based framework to evaluate the number of taxis that are scheduled to travel from one region to the other in an occupied condition, i.e., with a passenger on board. Thirdly, we elaborated the vehicle rebalancing model to evaluate the number of taxis that are recommended to travel from one region to the other in an unoccupied condition, with an objective to maximize the future served demands. Finally, we presented the mechanism for CMoD fleet control. We used a Receding Horizon Control (RHC) to update the supply distribution, served demand ratio, profit and cost of operation. We tested the performance of the proposed CMoD model by testing its supply-demand balancing efficacy over the Manhattan market and the results show that using the proposed model leads to higher served demands in the near future.

Chapter 3

Learning Optimal Model Parameters

3.1 Learning the Optimal Parameters

3.1.1 Motivation

In this subsection, we discuss the impact of two parameters, i.e., the balancing frequency f and the horizon window size ω , to test the successful implementation of the vehicle rebalancing.

The operating area is divided into n regions and the current time slot is denoted as k . We look ahead ω time steps i.e., k to $k + \omega$. This duration is known as the prediction horizon. We anticipate the passenger pickup demands and the taxi supply that is en-route each region during this interval. The value of the prediction horizon is critical because short prediction horizon doesn't allow sufficient time for the vehicles to be re-balanced on the other hand very large prediction horizon leads to higher level of uncertainty in the prediction. Secondly, how often, do we need to perform the vehicle rebalancing is another important consideration. The reason being that performing vehicle rebalancing very frequently may cause the vehicles to be en-route one region to the other in an unoccupied condition. This may cause excessive fuel cost expenditures. Another question that arises is that whether, performing vehicle rebalancing is equally beneficial during each time slot of the day or does the passenger demand market influences the decision to perform the vehicle rebalancing.

The objective is to find the optimal value of parameters, i.e., the balancing frequency f and the prediction horizon window ω . This will improve the performance of the vehicle rebalancing policy.

In this chapter, we first elaborate the significance of the two model parameters, namely the horizon window ω and the balancing frequency f in the efficacy of the supply-demand balancing capacity of the model. Then we present an Upper Confidence Bound based algorithm to learn the optimal model parameters. Finally, case studies to analyze the impact of the model parameters.

Frequency of Vehicle Rebalancing (f)

The supply scheduling problem (2.6) is solved during each time slot in order to serve the current demands. However, whether we should perform the vehicle rebalancing in each time slot needs to be decided before hand. The reason is that, the vehicle rebalancing requires the vehicles to travel without passengers which incurs extra cost. So, unless the rebalancing causes substantial improvement in the performance of the fleet, it might be better for the vehicles to wait in their respective regions. So, we choose whether to solve the vehicle rebalancing problem (2.11) during time slot k using the decision variable $\delta(k) \in \{0, 1\}$ which is equal to 1 if vehicle rebalancing is planned for time slot k ; and 0 otherwise.

Let $\Delta = [\delta(1), \delta(2), \dots, \delta(k), \dots, \delta(K)]$ be a vector that stacks all these decision variables together. Frequency (f) is the ratio of the time slots in which vehicle rebalancing problem is solved to the total number of time slots K :

$$f = \frac{\sum_k \delta(k)}{K} \quad (3.1)$$

The following figure shows the vector Δ as a side bar. Each term $\delta(k) \in 0, 1$ decides if vehicle rebalancing is planned for time slot k . As seen in the figure, the vehicle rebalancing is planned for time slot k and $\delta(k) = 1$.

Figure 3.1: Frequency of Vehicle Rebalancing

In order to decide, in which time slot should we perform the vehicle rebalancing, i.e., determine the value of $\delta(k) \forall k$, we need to consider the impact of the demand markets. This is because, during the peak hours, there is a higher probability of a driver being scheduled however, we have less number of idle drivers at that time. On the other hand during off-peak hours, we have more number of idle standing drivers but less number of passenger requests. So, it is not trivial to decide, which time slots should be chosen for vehicle rebalancing, if the balancing frequency is fixed.

Let $\Gamma = [\gamma(1), \gamma(2) \dots \gamma(k), \dots \gamma(K)]$ be a preference vector, where $\gamma(k)$ is the ratio of the expected demands in the time slot k to the overall expected demand:

$$\gamma(k) = \frac{\sum_{ij} [a_{ij}(k)]}{\sum_k \sum_{ij} [a_{ij}(k)]} \quad (3.2)$$

We used two extreme cases to check the impact of the balancing frequency on the vehicle rebalancing performance. For a fixed value of f , and in a given market condition. First, we fixed the value of the balancing frequency. Then, we used the following two case studies to determine, vehicle rebalancing should be performed in which time slots.

- Case 1: We choose the time slots for which the passenger pickup demands are very high, i.e., we choose the peak hours for vehicle rebalancing. The value of vector Δ is obtained by solving the following optimization problem. The motivation behind solving the problem (3.3) is to ensure that for a fixed value of f , the time slots with higher expected demand values are given a higher preference while selecting the time slots for the vehicle rebalancing.

$$\begin{aligned} \max_{\Delta} \quad & \sum_k \gamma(k) \cdot \delta(k) \\ \text{s.t.} \quad & \sum_k \delta(k) = f \cdot K \end{aligned} \quad (3.3)$$

- Case 2: We choose the time slots for which the passenger pickup demands are very low, i.e., we choose the off peak hours for vehicle rebalancing. The value of vector Δ is obtained by solving the following optimization problem. The motivation behind solving the problem (3.4) is to ensure that for a fixed value of f , the time slots with lower expected demand values are given a higher preference while selecting the time slots for the vehicle rebalancing.

$$\begin{aligned} \min_{\Delta} \quad & \sum_k \gamma(k) \cdot \delta(k) \\ \text{s.t.} \quad & \sum_k \delta(k) = f \cdot K \end{aligned} \quad (3.4)$$

In the subsection 3.3.1, we discuss the impact of the balancing frequency, in a given demand market setting, on the vehicle rebalancing, policy.

Horizon Window (ω)

While forecasting the demand or supply in each zone j , we need to decide how far in the future should we look ahead. This is termed as the prediction horizon window, which is denoted by ω . The choice of the appropriate horizon window is important because we suspect that too short horizon window is prone to myopic vision and hence, leads to lower served demands. The reason being that the vehicles that are dispatched from one region to the other should be able to reach the destination within the prediction horizon window. So, if the window is kept too short, then only few taxis can be dispatched from one region to the other. Due to this, we may not be able to get sufficient advantage of the vehicle rebalancing strategy. On the other hand too large prediction horizon may also lead to a drop in the performance of the vehicle rebalancing strategy because of increase in the uncertainty in the anticipated demand and supply.

We use a range of horizon window ω values to test the performance of the model and discuss the results in section 3.3.1

Figure 3.2: Horizon Window

3.1.2 Methodology

The objective is to learn the optimal set of parameter values, i.e., critical model parameters namely the balancing frequency f and the horizon window ω . We use a single state Markov Decision Process to learn the optimal set of model parameters. The state of the system is the initial distribution of the vehicles, i.e., the initial number of vehicles in each zone.

Starting in this state, an *action* is picked using the Upper-Confidence-Bound (UCB) action selection method. The action corresponds to a tuple of model parameter values, i.e., $a = \{f, \omega\}$ that decides the frequency of vehicle rebalancing (f) and horizon window (ω) for a given noise parameter of our predictor.

Using action $a = A_t$ as the input, the problems (2.6), (2.11) are solved in a sequential manner all the way till the last time-slot of the day thus concluding the episode with the action-value $r(s, a)$. In our experiment, we have chosen the Served Demand Ratio (*SDR*) as the action value. We update the action value pertaining to the action A_t using the observed reward $r(s, A_t)$, this updated value influences the action selection in the future, i.e., the actions with the higher action value are given the preference. Apart from this we also need to explore the actions for which the reward has not been observed yet. Hence, Upper Confidence Bound action selection method was used, which gives an ideal exploration-exploitation mixture.

Fig 3.3 shows the flow of this process. We always start with state s , choose the action using UCB action selection, observe the reward, update the action value and this process is repeated until optimal set of parameters are obtained. The details of the algorithm are explained next.

Figure 3.3: Single Step Markov Decision Process

3.1.3 Algorithm

The objective is to learn the action which has the highest action-value $Q(a)$ which is done as follows:

- Horizon window ω and frequency f are sampled as a tuple from uniform distributions say, $\omega \in U[0,60]$, i.e., 0 to 60 mins and $f \in U[0,1]$ respectively. The reason for choosing a uniform distribution is that this distribution helps in effective exploration over the defined set of possible values, i.e., we get the values from all over the defined range.
- Preference vector Γ is generated using the sampled value of f . The preference vector which is calculated using the problem 3.3, which decides which time slot should be used for vehicle rebalancing.
- A set of tuples $\{\omega, f\}$ form the action space Ω such that $a = \{\omega, f\} \in \Omega$. We need to find the optimal action out of all the actions in the set Ω .
- Each action $a \in \Omega$ is initialized with an action-value $Q(a) = 0$ because, we don't know the value of choosing any action yet.
- An action $A_t \in \Omega$ is chosen using the UCB action-selection. The reason of choosing the UCB policy is that it leads to an effective exploitation-exploration combination. The UCB action selection is as follows:

$$A_t = \underset{a}{\operatorname{argmax}} \left[Q_t(a) + c \sqrt{\frac{\ln(t)}{N_t(a)}} \right] \quad (3.5)$$

The term $Q_t(a)$ which is the action value for action a , favours the selection of action with the highest value (exploitation) and the term $\sqrt{\frac{\ln(t)}{N_t(a)}}$ favours the selection of action that have been selected less number of times, which is recorded in $N_t(a)$ to be preferred (exploration). Thus, the over all effect is to perform exploration and exploitation. The hyper parameter c , decides the importance that should be given to exploration.

- Using the action A_t , problems (2.6), (2.11) are solved sequentially till the episode terminates with a reward $r(A_t) = SDR$, i.e., the value of reward of choosing the action A_t is equal to the served demand ratio. The action value is updated as follows:

$$Q_t(a) := Q_t(a) + \frac{1}{t} [r(A_t) - Q_t(a)] \quad (3.6)$$

- Iteration index is incremented and the process repeats till the action values donot change significantly.

$$t := t + 1$$

3.2 Impact of Supply-Demand Balancing Policy

As a part of our analysis, two operational policies were compared in order to gauge the performance of the proposed policy.

- The first policy is called the waiting policy. In this operational policy, the idle standing drivers wait in their respective taxi zones after serving the demand.
- In the second case, i.e., dynamic supply-demand balancing policy, in which apart from supply scheduling (3), we also perform the vehicle rebalancing (4), whose optimal parameters, i.e., $\{\beta, \omega\}$ for a pre-defined noise-parameter α are learnt as explained in the previous section.

3.2.1 Impact of Implementing Vehicle Rebalancing

Fig 3.4 shows the overall city-wide passenger demand pattern. This was extracted using NYC Yellow Taxi Data [8]. There are two peaks, i.e., in the morning and the night corresponding to the office hours. In section 3.3, we discuss the impact of the proposed vehicle rebalancing model and Fig 3.6 shows the supply-demand imbalance mitigation capability corresponding the the morning peak, i.e., 6:00 – 10:00 hours.

Figure 3.4: Passenger Demand Pattern (Manhattan Jan, 20 2020)

Learning Optimal Parameter Set

In-order to test the impact of the proposed policy, all the zones $j \in Z$ were always initialized with same number of taxis, such that the number of taxis in any zone was proportional to the expected demand in that zone.

Then optimal set of parameters was found using the dynamic supply-demand balancing policy using the methodology as explained in previous section.

As shown in Fig 3.5, the episode reward increases as the policy parameters approach their optimal values. Once the optimal parameters are found, the episode reward converges to a constant value.

To quantify the advantage of using the proposed policy, *SDR* was recorded in both the case studies.

It was observed that using the optimal set of parameters denoted by $\{f^*, \omega^*\}$ that was found using the the proposed policy, lead to a 150% increase in the *SDR* as compared to when the *waiting policy*, i.e., $\{f = 0, \omega\} \in \Omega$ was used.

Figure 3.5: Learning Optimal Set of Parameters

3.2.2 Case Studies

In this sub-section, we present the results of three case studies that show the importance of choosing an appropriate set of parameters.

Horizon Window

Horizon window is the number of time slots ahead of the current time slot, that are used to predict the passenger demand and solve the supply-demand balancing optimization problem 2.11 that involves vehicle rebalancing.

As seen in Fig 3.6, the performance (SDR) of the supply-demand balancing model increases with the horizon window initially, reaches its peak, and decreases again. We observe this behavior as a result of the two contrasting forces at play:

- Increasing ω leads to a better vision about the upcoming supply-demand gap. This leads to higher served-demands.
- Increasing ω also leads to sending taxis to a very far away zone due to an increase in reach-ability index. These far-away recommendations are prone to higher levels of demand uncertainty due to drop-outs and prediction inaccuracy. Also, the taxis spend alot of the time traveling in unoccupied conditions leading to lower SDR.

Figure 3.6: Horizon Window Analysis

Analysis

In this subsection, we will provide the mathematical justification of the horizon window analysis as shown in the Fig 3.6. Recall that the vehicle rebalancing model we have used is as follows:

$$\begin{aligned}
 \max_{z,y} \quad & \sum_{j=1}^n z_j - \beta \sum_{i,j=1}^n c_{ij} y_{ij}(k) \\
 \text{s.t.} \quad & \sum_{j=1}^n y_{ij}(k) \leq v_i(k) \quad \forall i \in Z \\
 & (t_{ij} - \omega) \cdot y_{ij}(k) \leq 0 \quad \forall i, j \in Z \\
 & z_j \leq s_j(k, \omega) + \sum_i y_{ij}(k) \quad \forall j \in Z \\
 & z_j \leq \widetilde{\varphi}_j(k, \omega, \alpha) \quad \forall j \in Z
 \end{aligned}$$

Reachability index $\varphi = \frac{1}{n^2} \sum_{i,j|i \neq j} \rho_{ij}$ where, $\rho_{ij} = 1$ if $t_{ij} \leq \omega$ and $\rho_{ij} = 0$ if $t_{ij} > \omega$.

Predicted demand $\widetilde{\varphi}_j(k, \omega, \alpha) = \sum_{t=k+1}^{k+\omega} (\sum_i a_{ji}(t) + N(0, g(t)))$ where, $g(t) = \alpha t$,

Consider the following three cases:

- Short horizon window: $\varphi = 0$
 To satisfy the constraint 2 in the vehicle rebalancing problem, i.e., $(t_{ij} - \omega) \cdot y_{ij}(k) \leq 0$, since, $t_{ij} \leq \omega \forall i, j \in Z \mid i \neq j$, y_{ij} must be equal to 0.
 Therefore, there can be no rebalancing and hence, the drivers have to wait in their respective regions. Due to this, the benefits of the vehicle rebalancing cannot be achieved and the SDR is low.
- Long horizon window: $\varphi = 1$
 In this case, it's possible to travel between any two regions within the prediction horizon. The SDR drops due to the following two reasons, First, the taxis can be sent to very far away regions since, $\varphi = 1$. Therefore, the taxis spend a lot of time travelling to the far away zones which leaves lesser time for scheduling. Secondly, due to increase in ω , the error in the predicted demand, i.e., $\widetilde{\varphi}_j(k, \omega, \alpha)$ increases. this leads to poor recommendations and hence a lower SDR.
- Medium horizon window: $0 \leq \varphi \leq 1$
 In this case, there are two contrasting effects at play. If we consider the constraint 2, i.e., $(t_{ij} - \omega) \cdot y_{ij}(k) \leq 0$, as we increase ω , we are able to dispatch the vehicles from one region to the other. Hence, we start observing the benefits of the vehicle rebalancing. However, if we keep on increasing the value of ω , the effects of long horizon window as explained above, start to dominate. So we see an initial rise, and subsequent fall in the SDR as we increase the horizon window ω .

Balancing Frequency

Balancing frequency decides how often do we engage in the vehicle rebalancing strategy.

As seen in the Fig 3.7, increasing the frequency of vehicle rebalancing leads to a contrasting performance outcome with regards to the *SDR* and recommendation efficiency (η).

The Served Demand Ratio increases with the increase in the balancing frequency which means that incessant vehicle rebalancing leads to higher served demands due to excessive rebalancing within nearby regions (if ω is chosen appropriately).

However, we observed that continuous vehicle rebalancing leads to a drop in the recommendation efficiency that leads to higher operational costs, which may not be desirable.

Hence, continuous vehicle rebalancing in every time slot is not an optimal strategy when the objective of the company is to ensure a balance between the served demands and the operational costs.

In order to ensure a balance between the served demands and the operational costs the reward function for the supply-demand balancing policy must also include the operational cost while searching for the optimal model parameters.

Figure 3.7: Frequency Analysis

Impact of Demand Markets

For a fixed value of the frequency (f) of vehicle rebalancing and the horizon window (ω), we need to decide which time slots should be selected for implementing the vehicle rebalancing. In order to find these time slots, we compare two extreme scenarios. For both the cases, the frequency (f) of vehicle rebalancing and the horizon window (ω) is kept the same. The two cases are as follows:

- In the first case as shown in Fig 3.8, the time slots selected for supply-demand balancing correspond to the peak hours, i.e., the morning and the evening peak. For getting the time slots corresponding to the peak hours we solve the problem 3.3. More specifically, we calculate the value of vector Δ that decides the time slots in which vehicle rebalancing should be planned.
- In the second case, the chosen time slots are from the off-peak hours. For getting the time slots corresponding to the off-peak hours we solve the problem 3.4.

It was observed that the case 1 lead to 54% higher SDR and 13% higher recommendation efficiency (η).

Figure 3.8: Demand Market Analysis

Hence, the selection of the time slots for vehicle rebalancing depends on the demand market and it is desirable to prefer the peak hours for the rebalancing.

Analysis

In this subsection, we will provide the mathematical justification of the horizon window analysis as shown in the Fig 3.7 and 3.8. Recall that the vehicle rebalancing model that we have used is as follows:

$$\begin{aligned}
 \max_{z,y} \quad & \sum_{j=1}^n z_j - \beta \sum_{i,j=1}^n c_{ij} y_{ij}(k) \\
 \text{s.t.} \quad & \sum_{j=1}^n y_{ij}(k) \leq v_i(k) \quad \forall i \in Z \\
 & (t_{ij} - \omega) \cdot y_{ij}(k) \leq 0 \quad \forall i, j \in Z \\
 & z_j \leq s_j(k, \omega) + \sum_i y_{ij}(k) \quad \forall j \in Z \\
 & z_j \leq \widetilde{\phi}_j(k, \omega, \alpha) \quad \forall j \in Z
 \end{aligned}$$

Recommendation efficiency η which is given by $\eta = \frac{\sum_k \sum_j ((\sum_{t=k+1}^{k+\omega} \sum_i x_{ji}(t)) - s_j(k, \omega))}{\sum_k \sum_{i,j} y_{ij}(k)}$.

It was observed that during the off-peak hour, the number of recommendations that were made (case 2), were significantly lower than the number of the recommendations that were made during the peak hours (case 1).

This is because of the combined effect of the constraint 3,4 and the second term of the objective function ($-\beta \sum_{i,j=1}^n c_{ij} y_{ij}(k)$) that ensures minimization of the idle mileage.

According to the constraints 3 and 4, the anticipated served demand is less than the predicted demand (constraint 4) and the supply (constraint 3). Since, during the off-peak hours the demand is very low, so to satisfy the constraint 4, the anticipated served demand (z_j) also has to take a low value. Since, we are minimizing the idle mileage ($-\beta \sum_{i,j=1}^n c_{ij} y_{ij}(k)$), the number of vehicles that can be rebalanced (y_{ij}) which also forms the part of the constraint 3 must be low as well.

Due to the lower number of recommendations, the benefits of vehicle rebalancing cannot be gained during the off-peak hours. Hence, the SDR is lower when off-peak hours are selected for performing vehicle rebalancing.

It was also observed that incessantly increasing the balancing frequency led to a degradation in the recommendation efficiency as seen in the figure 3.8. This is because with an increase in the balancing frequency, the number of recommendations ($\sum_k \sum_{i,j} y_{ij}(k)$), which is the denominator in the recommendation efficiency calculation is going to increase. However, the total number of useful recommendations which is the numerator may not increase in the same proportion, the reason being that the number of served demands ($\sum_{i,j,k} x_{ij}(k)$) depends on factors like the demand market as well as the accuracy of the demand predictor, etc.

Impact of Demand Prediction Inaccuracies

Since the demand prediction is prone to uncertainties, in this section we demonstrate the impact of the demand prediction inaccuracies on the performance of the proposed model. The inaccuracies in the predicted demand impacts the model due to the constraint 4, i.e.,

$$z_j \leq \widetilde{\rho}_j(k, \omega, \alpha) \quad \forall j \in Z$$

Here, $\widetilde{\rho}_j(k, \omega, \alpha)$ is the anticipated demand, which is a function of the prediction horizon ω as well as the noise factor α . Recall, the equation for the predicted demand is as follows:

$$\widetilde{\rho}_j(k, \omega, \alpha) = \sum_{t=k+1}^{k+\omega} \left(\sum_i a_{ji}(t) + N(0, g(t)) \right)$$

Where,

- $a_{ji}(t)$: actual demand to go from region $j \rightarrow i$
- $N(0, g(t))$: gaussian noise

The Gaussian noise with a zero mean and standard deviation which is linear in time, i.e., $g(t) = \alpha t$ was used. in this experiment the impact of demand prediction inaccuracies was quantified in terms of reduction in the Served Demand Ratio, SDR and the results are as follows:

Figure 3.9: Impact of Demand Prediction Inaccuracies

As seen in Fig 3.9, the SDR reduces as the value of α increase. This means that the performance of the model degrades with the increase in demand prediction inaccuracies. Therefore having an accurate demand predictor plays a crucial role in the performance of the model.

3.3 Conclusion

We presented the proposed coordinated mobility-on-demand system model. The model has three primary components, namely the supply-scheduling system, supply-demand balancing system (vehicle rebalancing) and the fleet control mechanism. We discussed the detailed mathematical formulation and key components of our model and assessed the model parameters that are critical to ensure the supply-demand balancing efficacy. Firstly, we analyzed the frequency at which the balancing should be done, i.e., how often and during which time of the day do we need to perform the vehicle rebalancing. We showed through the realistic simulations that it is neither necessary nor desirable to continuously rebalance the taxis throughout the day. More specifically, the under-utilization of the rebalancing strategy leads to less capacity utilization and lower served demands whereas, incessant rebalancing leads to significant degradation of the recommendation efficiency which leads to a higher cost of operation to the company. We also showed the advantages of choosing the peak demand hours for vehicle rebalancing over non-peak hours. Secondly, we analyzed the impact of the prediction horizon. A phenomenon, similar to the Wild Goose Chase was observed for large prediction horizons. The served demand showed an initial rise w.r.t increase in the horizon window, however, the performance started to degrade for the larger prediction horizon. We observed a high sensitivity in the system performance with regards to perturbations in these parameters. Finally, we presented a methodology to learn the optimal set of parameters using the proposed dynamic supply-demand balancing policy and it was observed that the optimal model parameters can lead upto 150% improvement in supply-demand balancing.

Chapter 4

Ensuring Service Fairness in CMoD Model

4.1 Motivations and Problem Statement

In a Coordinated Mobility on Demand (CMoD) system, the taxi drivers follow the instructions of a central scheduling and recommendation system. When a taxi driver is standing idle, the system recommends him a strategy to find his next passenger. The system may recommend a taxi driver to stay at the same location and wait for the next request to show up or it may recommend him to go to some other location where his chance of finding a passenger is higher. The objective of these kind of systems is generally to maximize the company profit or minimize the passenger waiting time and the decisions are made from a higher level of abstraction as explained in the hierarchy of the CMoD. Recall, that the proposed CMoD model had a two level hierarchy. At the top level, the operational decisions like the supply scheduling and the vehicle rebalancing are implemented where we decide how many taxis should be dispatched from one region to the other. At the lower level, we decide which driver should be scheduled and which driver should be recommended to the other regions. This decision is made taking various factors into account such as the driver profits, whether he has recently served a recommendation, etc and is termed as Service Fairness.

In this chapter, we have integrated supply scheduling, vehicle rebalancing and service fairness into a comprehensive CMoD fleet management model. By doing so, we have addressed one of the challenges in implementing a driver-centered coordinated taxi fleet system, which is the driver satisfaction. The primary motivation behind this work is that, we believe the taxi drivers shall adhere to the system recommendations (vehicle rebalancing) which involves travelling idle distance, only if they see a clear benefit like increased profits. From a driver's perspective, he expects to get his next ride as soon as he reaches the recommended location. We term this assurance as the scheduling guarantee. Secondly, the drivers must feel that the system is fair. This can be implemented by reducing the disparity among the drivers' profits. These two components have been

used to quantify Service Fairness. We upgrade the CMoD model that was presented in previous chapters to incorporate service fairness factor into the model thus integrating both the top and bottom level operations of the CMoD hierarchy into a single framework. This unified framework helps in achieving the service fairness while simultaneously ensuring higher level of company profits due to a better level of control over the supply distribution.

Consider the following example, As seen in the Fig 4.1, we have two taxis standing idle in region 1. One of the taxi has already earned 100\$ till now and the second taxi has just earned 10\$. Also, in region 2 we see that one of the taxi has just served a recommendation ($r = 1$), i.e., travelled from other region to region 2 whereas, the other taxi was already standing idle ($r = 0$).

Now, it depends how one views fairness. One such way can be that we prefer the drivers which are lagging in the profits or have recently served a recommendation being preferred for scheduling as shown in the figure below. The objective thus is to build a framework that provides a flexibility to incorporate the service fairness objectives.

Figure 4.1: Service Fairness

4.2 Mathematical Formulation

In each time-slot, we seek the best possible taxi dispatch solution. We need to decide which taxi should serve the current demand in the region that they are standing and which taxis should be recommended to go to other regions or wait in the same region to balance the future supply and demand. While doing so, we need to ensure the service fairness, which we quantify using two indicators namely the driver profit disparity and scheduling guarantee. While generating the dispatch solutions, we need to ensure that profit disparity among the drivers is reduced, i.e., drivers that are lagging in the profits should be given the preference to get scheduled. We also give the preference to the drivers, who have recently served a recommendation to get scheduled. This helps in increasing the scheduling guarantee, i.e., the driver who is serving a recommendation has a higher chance of getting scheduled when he reaches the destination.

4.2.1 Fleet Management Objectives:

Coordinated taxi fleet management has been formulated as a multi-objective constrained optimization problem. The multiple objectives that we need to serve are as follows:

- To serve the current demands: This is the primary objective from a company's point of view. We need to serve the current demands in each region and hence schedule as many taxis as possible to go from one region to the other in an occupied condition.
- To perform vehicle rebalancing: We need to check if we should dispatch the surplus taxis in each region to the other regions such that maximum number of demands can be served in the near future. The taxis are surplus in case when the number of idle standing taxis are more than the number of current passenger requests. For vehicle rebalancing, we need to anticipate the demands and supplies in each region in the near future and redistribute the surplus taxis among the regions.
- To minimize the average unloaded travel distance: While performing the vehicle rebalancing, we need to ensure that the overall distance the taxis in the fleet will have to travel in the unoccupied condition, which is referred to as the unloaded travel distance, is minimized.
- To ensure service fairness: We need to ensure that the taxis that are lagging in the profits earned till the current time slot and the taxis that have recently served a recommendation, i.e., were dispatched to the region in an unoccupied condition gets scheduled with priority. Giving priority to the drivers that are lagging in the profits reduces the profit disparity among the drivers and giving the priority to the drivers that have recently served the recommendation increases the scheduling guarantee.

4.2.2 Decision Variable

Variables $x(k)$, $y(k)$ and $z(k)$ together represent a dispatch solution at time slot k for a coordinated taxi fleet management problem.

- Supply scheduling variable $x(k)$:
 $x_{ijc}(k) \in \{0,1\}$. If taxi $c \in C$ has been scheduled to serve a demand during time slot k to go from region $i \in Z$ to $j \in Z$ then $x_{ijc}(k) = 1$ otherwise $x_{ijc}(k) = 0$.
- Vehicle rebalancing variable $y(k)$:
 $y_{ijc}(k) \in \{0,1\}$. If an idle taxi $c \in C$ has been recommended during time slot k to go from region i to j without any passenger, i.e., in order to balance the future supply-demand in various regions, then $y_{ijc}(k) = 1$. Taxis can also be recommended to stay in same region i.e. $j = i$.
- Future served demand variable $z(k)$:
 $z_j(k) \in R$ is the number of requests that can be served in region $j \in Z$ during the prediction horizon k to $k + \omega$.

Figure 4.2: Multi-Objectives of Taxi Fleet Management

4.2.3 Cost Function

Summarizing the aforementioned four objectives, our dispatch solution $x(k)$ and $y(k)$ should be such that the cost function $J(k)$ is maximized during each time slot k :

$$\max_{x(k), y(k)} J(k) = S(k) + E(k) + ID(k) + F(k) \quad (4.1)$$

- Served Demand $S(k)$:

The demand, i.e., the number of passenger requests to go from region $i \in Z$ to another region $j \in Z$ in time slot k is denoted as $a_{ij}(k)$. $S(k)$ in any time slot k is equal to the total number of passenger requests that will be served in the entire city.

The Served Demand $S(k)$ is captured by the following expression:

$$S(k) = \sum_{c=1}^N \sum_{i=1}^n \sum_{j=1}^n x_{ijc}(k) \quad (4.2)$$

- Expected Future Served Demand $E(k)$:

If we can predict the future demand $\widetilde{\wp}_j(k, \omega, \alpha)$, i.e., the number of requests in each region j for a prediction horizon ω and the number of vehicles $s_j(k, \omega)$ that will be available in each region j , we can estimate the demand which can be served.

The vehicle supply in each region j in next time slot is given by:

$$s_j(k, \omega) = \sum_{c=1}^N \sum_{i=1}^n y_{ijc}(k) + \sum_{t=k+1}^{k+\omega} \sum_{c=1}^N v_{jc}(t) \quad (4.3)$$

The expected number of passenger request for each region j in next time slot is as follows:

$$\widetilde{\wp}_j(k, \omega, \alpha) = \sum_{t=k+1}^{k+\omega} \left(\sum_i a_{ji}(t) + N(0, g(t)) \right) \quad (4.4)$$

The expected demand that will be served is denoted by $E(k)$ and is captured using the following equation:

$$E(k) = \gamma \sum_{j=1}^n z_j(k) \quad (4.5)$$

Where,

$$z_j(k) = \min(s_j(k, \omega), \widetilde{\wp}_j(k, \omega, \alpha)) \quad (4.6)$$

- Service Fairness $F(k)$: Service Fairness has been quantified as scheduling guarantee and driver profit disparity.

Scheduling guarantee means that the driver who was recommended to go to other location in the previous time-slot will have a higher chance of getting scheduled in the current time-slot when he reaches that location.

In order to implement the scheduling guarantee, we record if a taxi is recommended to go to the other region during each time slot k in a vector $p(k)$, where for taxi $c \in C$

$$p_c(k) = \sum_{j=1}^n \sum_{i=1}^n y_{ijc}(k) - \sum_{i=1}^n y_{iic}(k) \quad (4.7)$$

Secondly the taxis that are lagging on the profits are given a priority in the scheduling. To reduce the profit disparity, we record profits of individual taxi drivers in a vector $P(k)$, where $P_c(k) \forall c \in C$ is the profit that taxi c has made so far. A normalized profit vector is given as:

$$\begin{aligned} P_c^N(k) &= P_c(k) / \max(P(k)) \\ P_c^{N'}(k) &= 1 - P_c^N(k) \end{aligned} \quad (4.8)$$

Where, $P^{N'}(k)$ is the inverse normalized profit disparity vector.

The Service Fairness $F(k)$ is captured using the following equation:

$$F(k) = \eta * \sum_{c=1}^N ((P_c^{N'}(k) + p_c(k-1)) * \sum_{j=1}^n \sum_{i=1}^n x_{ijc}(k)) \quad (4.9)$$

- Idle Distance $ID(k)$: Idle distance that the taxis will have to travel is minimized, which is captured by maximizing the following expression, where c_{ij} is the cost of travelling from region $i \in Z$ to $j \in Z$.

$$ID(k) = -\beta * \sum_{c=1}^N \left(\sum_{j=1}^n \sum_{i=1}^n y_{ijc}(k) * c_{ij} \right) \quad (4.10)$$

Taxi Fleet Management Component Summary:

- Demand Matrix $A(k)$: $a_{ij}(k) \in R_+$ is the number of passenger requests in time slot k from region $i \in Z$ to region $j \in Z$.
- Supply Matrix $V(k)$: $v_{ic}(k) \in \{0,1\}$ is equal to 1 iff taxi $c \in C$ is standing idle in region $i \in Z$.
- Predicted Supply $s(k, \omega)$: $s_j(k, \omega) \in R$ is the predicted supply, i.e., the number of taxis that shall be available in region $j \in Z$ during time period k to $k + \omega$.
- Predicted Demand $\tilde{\phi}(k, \omega, \alpha)$: $\tilde{\phi}_j(k, \omega, \alpha) \in R$ is the predicted demand, i.e., the number of passenger requests that will appear in region $j \in Z$ during time period k to $k + \omega$.
- Cost Matrix C : $c_{ij} \in R_+$ is the fuel cost of travelling from region $i \in Z$ to $j \in Z$.
- Inverse Normalized Profit Matrix $P^{N'}(k)$: $P_c^{N'}(k)$ is related to the profit earned by the taxi $c \in C$ in the following manner:

$$P_c^{N'}(k) = 1 - P_c(k)/\max(P(k))$$

- Recommendation Matrix $p(k)$: $p_c(k) \in \{0,1\}$ is equal to 1 iff taxi $c \in C$ was recommended to go from one region to the other during time slot k .
- Travel Time Matrix T : $t_{ij} \in T$ is the time that is taken on average to travel from the region $i \in Z$ to $j \in Z$.

4.2.4 Constraints:

1. The served demand in each region $i, j \in Z$ should be less than the actual demand during every time slot k . Since, $x_{ijc}(k)$ decides if the taxi $c \in C$ is dispatched from region $i \in Z$ to $j \in Z$ in an occupied condition, i.e., the taxi $c \in C$ is scheduled to go from region i to go to region j . The term $\sum_{c=1}^N x_{ijc}(k)$ is the total number of taxis that are scheduled from region i to j . The total number of requests from region i to j , where the trip started in time slot k are recorded in the demand matrix and is denoted as $a_{ij}(k) \in A(k)$. The total number of scheduled taxis is less than the total number of passenger requests from region i to j .

$$\sum_{c=1}^N x_{ijc}(k) \leq a_{ij}(k) \quad \forall i, j \in Z \quad (4.11)$$

2. A vacant taxi $c \in C$ should either be scheduled or recommended at the end of each time slot, i.e., the for all the taxis $c \in C$ that are standing idle in region $i \in Z$ should either be scheduled i.e., dispatched with a passenger on board or should be recommended a region to wait for his next ride. The recommended region can be the same region in which he is already standing. In the equation below, the expression on the right tells if taxi $c \in C$ is standing idle in any region $i \in C$. Note, if taxi c is standing idle in region i then $v_i(k) = 1$, else it is 0. So, if the expression on the right is 1 then expression on the left has to be equal to 1, i.e., either the taxi should be scheduled where, $\sum_{j=1}^n \sum_{i=1}^n x_{ijc}(k) = 1$ or the taxi should be recommended to go from region i to j . In that case $\sum_{j=1}^n \sum_{i=1}^n y_{ijc}(k) = 1$. The expression on the left also ensures that the taxi is either scheduled or recommended.

$$\sum_{j=1}^n \sum_{i=1}^n (x_{ijc}(k) + y_{ijc}(k)) = \sum_{i=1}^n v_{ic}(k) \quad \forall c \in C \quad (4.12)$$

3. A vacant taxi $c \in C$ is scheduled or recommended from the same region $i \in Z$ in which it is currently standing vacant. The expression on the right i.e., $v_{ic}(k) = 1$ if taxi $c \in C$ is standing idle in region $i \in Z$ otherwise it is 0. So, if the taxis $c \in C$ is standing idle in region $i \in Z$, the taxi must be either scheduled or recommended to region $j \in Z$.

$$\sum_{j=1}^n (x_{ijc}(k) + y_{ijc}(k)) = v_{ic}(k) \quad \forall i \in Z, c \in C \quad (4.13)$$

4. Variable $z_j(k) = \min(s_j(k, \omega), \widetilde{\varphi}_j(k, \omega, \alpha))$ is used as a decision variable by introducing the following two constraints using (4.3) and (4.4). Equation 4.3 says that the predicted supply in each region $j \in Z$ is equal to the number of taxis that are already en-route region $j \in Z$ plus the number of taxis that will be recommended to go to region $j \in Z$ from all the regions $i \in Z$.

Since the anticipated served demands in region $j \in Z$ during the prediction horizon k to $k + \omega$ is less than or equal to the number of taxis that will be available in region $j \in Z$. We have the following equation explaining this idea:

$$z_j(k) \leq \sum_{c=1}^N \sum_{i=1}^n y_{ijc}(k) + \sum_{t=k+1}^{k+\omega} \sum_{c=1}^N v_{jc}(t) \quad \forall j \in Z \quad (4.14)$$

The anticipated served demands $j \in Z$ during the prediction horizon k to $k + \omega$ is less than or equal to the total number of passenger requests that are expected to appear during the prediction horizon. The predicted demand in region $j \in Z$ is shown on the right hand side of the equation 4.15. The calculation of the predicted demand has been explained in great detail in the last chapter. However, in short, we use the actual demand with a Gaussian noise added on top of that to model the predicted demand.

$$z_j(k) \leq \sum_{t=k+1}^{k+\omega} \left(\sum_i a_{ji}(t) + N(0, g(t)) \right) \quad \forall j \in Z \quad (4.15)$$

5. Taxi c should only be recommended to a location if it can reach there within the prediction horizon. As shown in the equation below, if the term $y_{ijc}(k) = 1$ and the constraint enforces $y_{ijc}(k) * t_{ij} \leq \omega$, which means that the a taxi $c \in C$ can only be recommended to go from region i to j if it can reach there within the prediction horizon ω :

$$\sum_{j=1}^n \sum_{i=1}^n (y_{ijc}(k) * t_{ij}) \leq \omega \quad \forall c \in C \quad (4.16)$$

4.2.5 Multi-Objective Taxi Fleet Management Problem Summary

$$\begin{aligned}
 & \max_{x(k), y(k), z(k)} J(k) = S(k) + E(k) + F(k) + ID(k) \\
 & \text{s.t.} \quad \sum_{c=1}^N x_{ijc}(k) \leq a_{ij}(k) \quad \forall i, j \in Z \\
 & \quad \sum_{j=1}^n \sum_{i=1}^n (x_{ijc}(k) + y_{ijc}(k)) = \sum_{i=1}^n v_{ic}(k) \quad \forall c \in C \\
 & \quad \sum_{j=1}^n (x_{ijc}(k) + y_{ijc}(k)) = v_{ic}(k) \quad \forall i \in Z, c \in C \\
 & \quad z_j(k) \leq \sum_{c=1}^N \sum_{i=1}^n y_{ijc}(k) + \sum_{t=k+1}^{k+\omega} \sum_{c=1}^N v_{jc}(t) \quad \forall j \in Z \\
 & \quad \sum_{j=1}^n \sum_{i=1}^n (y_{ijc}(k) * t_{ij}) \leq \omega \quad \forall c \in C
 \end{aligned} \tag{4.17}$$

Where,

- $S(k) = \sum_{c=1}^N \sum_{i=1}^n \sum_{j=1}^n x_{ijc}(k)$
- $E(k) = \gamma \sum_{j=1}^n z_j(k)$
- $F(k) = \eta * \sum_{c=1}^N ((P_c^{N'}(k) + p_c(k-1)) * \sum_{j=1}^n \sum_{i=1}^n x_{ijc}(k))$
- $ID(k) = -\beta * \sum_{c=1}^N \left(\sum_{j=1}^n \sum_{i=1}^n y_{ijc}(k) * c_{ij} \right)$

Figure 4.3: Multi-Objective Taxi Fleet Management Problem

4.3 A Receding Horizon Control Strategy

We apply a receding horizon strategy, when carrying out optimization in an iterative manner. During each decision horizon, we solve an optimization problem, and execute the decision values $x(k)$ and $y(k)$ for the first time interval. Then based on the actual system state at the end of the first time interval, we solve again the optimization problem, and the process repeats.

To ensure a success of this strategy, we need to continuously update model parameters, e.g., taxi locations and their status, which are described below.

- Taxi Locations:

Decision variables $x(k)$ and $y(k)$ tell the final location of the taxi and t_{ij} tells when the taxi will reach that location. We assume constant travel time t_{ij} for trip (i, j) during any time slot.

We update the vehicle location matrix v as follows:

$$v_{jc}(k') = 1 \quad (4.18)$$

Where,

$$k' = k + t_{ij} \quad (4.19)$$

- Driver Profit:

Vector $P(k)$ stores the profit that has been earned by each taxi $c \in C$ till the present time slot k . When a taxi completes a ride from region i to j , it earns $r_{ij} = m_{ij} - c_{ij}$ profit. After each iteration, the profit $P_c(k)$ is updated by:

$$P_c(k) = P_c(k-1) + \Delta P_c(k) \quad (4.20)$$

where

$$\Delta P_c(k) = \sum_{j=1}^n \sum_{i=1}^n (x_{ijc}(k) * (m_{ij} - c_{ij}) - y_{ijc}(k) * c_{ij}) \quad (4.21)$$

- Company Profit:

Vector $M_c(k)$ stores the money that has been earned by the company in time slot k . When a taxi completes a ride, the company that runs the on-demand mobility service gets a fraction of this amount. We denote this fraction as κ . So, after each time step, the profit of company is updated as follows:

$$M_c(k) = \kappa * \sum_{c=1}^N \sum_{j=1}^n \sum_{i=1}^n x_{ijc}(k) * m_{ij} \quad (4.22)$$

- Served Demand:

The number of requests that have been served in each time slot k is stored in a vector $\sigma(k)$. After each iteration the served demand is updated as follows:

$$\sigma(k) = \sum_{c=1}^N \sum_{j=1}^n \sum_{i=1}^n x_{ijc}(k) \quad (4.23)$$

The actual demand is updated as follows:

$$\Psi(k) = \sum_{j=1}^n \sum_{i=1}^n a_{ij}(k) \quad (4.24)$$

The following figure shows the strategy to control the CMoD fleet. During every time step k , we first solve the multi-objective constrained optimization problem for fleet management and generate a dispatch solution. We use the generated dispatch solution to update the state of the system, i.e. update the location of each taxi, driver profits, company profits and served demands.

Figure 4.4: Coordinated Mobility-on-Demand System Fleet Control

4.4 Simulation Case Studies

The supply-scheduling and vehicle rebalancing problem is modeled as a multi-objective taxi fleet management problem 4.17. The cost function consists of four components each corresponding to a separate objective. In order to test the impact of each cost function term, we created three scenarios of a coordinated taxi system. All three models used the same taxi network and constraints but with different cost functions.

- Model 1: $J(k) = S(k) + E(k) + ID(k) + F(k)$ The objective of this model was to maximize the current as well as the future demand that can be served while making sure that the idle distance that the taxis needed to travel was minimum. This model also maintained the service fairness among the drivers.
- Model 2: $J(k) = S(k) + E(k) + ID(k)$ The objective of this model was to maximize the current as well as future demand that can be served while making sure that the idle distance that the taxis need to travel was minimum.
- Model 3: $J(k) = S(k) + ID(k)$ The objective of this model was to only maximize the current demand that can be served with minimum idle distance. This replicated a system where no future demand prediction was employed in a fleet of coordinated taxis. The drivers always remained in their taxi zones with the objective of saving fuel

We carried out simulation case studies based on NYC Taxi Data [8] for a period of 30 days.

The Key Performance Indicators (KPIs) that helped us in gauging the improvements shown by the proposed models fall under two broad categories:

- Company Level Benefits: The company level benefits include the served demands and the company profits.
- Driver Level Benefits: The driver level benefits include the individual driver profits, driver profit disparity and the scheduling guarantee.

4.4.1 Company Benefits

The benefits at the company level can be quantified with two parameters namely the served demands and the company profits.

Served Demand

For each day $d \in [1, 30]$, we recorded the number of passenger requests served by End of Day (EoD). At each time slot $k \in [1, k_D]$, we recorded the served demand $\sigma(k)$. The total demand that got served EoD is given by $\sum_{k=1}^{k_D} \sigma(k)$ where, $x_{ijc}(k) \in \{0, 1\}$ if equal to 1 if taxi c is scheduled to travel from region i to j :

$$\sigma(k) = \sum_{c=1}^N \sum_{j=1}^n \sum_{i=1}^n x_{ijc}(k) \quad (4.25)$$

Fig 4.5 shows the comparison between the Actual Demand and EoD Served Demand for three models.

Figure 4.5: Daily served demand wrt actual demand for 100 Taxis

Company Profit

For each day $d \in [1, 30]$, we record the profits earned by the company EoD. At each time slot $k \in [1, k_D]$, we record the profit $M_c(k)$. For every dollar earned by the driver, the company keeps its fixed share say 20%.

The total profit that was earned by end of day EoD is given by $\sum_{k=1}^{k_D} M_c(k)$ where,

$$M_c(k) = \kappa * \sum_{c=1}^N \left(\sum_{j=1}^n \sum_{i=1}^n (x_{ijc}(k) * m_{ij}) \right) \quad (4.26)$$

Fig 4.6 shows the comparison between profits earned for three models. As seen, Model 1 and Model 2 performs equally well. However, Model 3 leads to a significantly lower profit value. The reason for this is that Model 3 doesn't deploy vehicle rebalancing.

Figure 4.6: Daily served demand wrt actual demand for 100 Taxis

4.4.2 Driver Benefits

To evaluate the performance of various Models in terms of driver benefits, we are using the following KPIs:

Mean Driver Profit

We recorded the daily profit of each driver which we call the EoD driver profit. We also record the End of Month (EoM) mean driver profit. The EoD mean profit of drivers is given by $P_{mean} = \frac{1}{N} \sum_{c=1}^N P_c(k_D)$ where,

$$P_c(k) = P_c(k-1) + \sum_{j=1}^n \sum_{i=1}^n (x_{ijc}(k) * (m_{ij} - c_{ij}) - y_{ijc}(k) * c_{ij})$$

Fig 4.7 shows the profit earned by each driver in a fleet of 100 taxis. As seen in the figure, the profit disparity for Model 1 is the lowest because of the service fairness term. Also, the mean driver profit for Model 1 and Model 2 was significantly higher than Model 3 because of vehicle rebalancing.

Figure 4.7: Profit earned by drivers at month end for different Models

Driver Profit Disparity

It might be possible that the mean value of a driver's profit is increasing but there is a lot of disparity among drivers' profit, i.e., some drivers are earning a lot more than others. A fair system should ensure minimum profit disparity among the drivers. We test this disparity by checking the standard deviation of driver profits:

$$P_{std} = \sqrt{\frac{\sum_{c=1}^N (P_c(k_D) - P_{mean})^2}{N - 1}} \quad (4.27)$$

Fig 4.8 is the daily record of standard deviation among the taxi driver profits over a period of 30 days for three Models. As seen in the figure, Model 1 leads to almost a constant value of the driver profit disparity as the profits are accumulated each day.

Figure 4.8: Profit disparity among drivers over a period of 30 days

Scheduling Guarantee

We keep a record of number of times the taxi was recommended to go to other regions labeled as the Total Recommendation Count (TRC). We also record how many times taxi $c \in C$ that was recommended to go to other region actually got scheduled within the prediction horizon ω , which is labeled as Scheduled Recommendation Count (SRC).

The metric that quantifies the Scheduling Guarantee is labelled as the Scheduling Guarantee Ratio (SGR). For taxi $c \in C$ the SGR is given as follows:

$$SGR(c) = \frac{SRC(c)}{TRC(c)} \quad (4.28)$$

Fig 4.9 compares the Scheduled Recommendation Count (SRC) vs Total Recommendation Count (TRC) for each taxi which is represented by a point in the scatter plot. In this figure, we have compared the SRC vs TRC for same fleet of 100 taxis, first while following the Model 1 and then Model 2.

Figure 4.9: Scheduled Recommendation Count Analysis

Discussion of Experimental Results

1. The $E(k)$ term in Model 1 and Model 2 objective function ensures supply-demand balancing in the prediction horizon and is responsible for the increased value of served demand and company profit.
 - Fig 4.5 shows using Model 1 leads to 49.84% higher served demand than Model 3 using 100 taxis and around 30% higher demand using Model 1 as compared to Model 3 when using 600 taxis.
 - It was observed that same level of demand can be served using Model 1 with $1/3^{rd}$ number of taxis as that of Model 3.
 - Fig 4.6 shows End of Month (EoM) company profit is 71 % higher for Model 1 and Model 2 than that of Model 3.
2. Service Fairness term $F(k)$ is responsible for reducing the profit disparity among the drivers as well as increasing the scheduling guarantee for the drivers who are serving the recommendations given by the system.
 - Fig 4.8 shows that for Model 1, we are able to contain the standard deviation of profits to a practically constant value. For a fleet of 100 taxis, the standard deviation in driver profits for Model 1 is 83% lower than that of Model 2 and Model 3.
 - The EoM mean driver profit for Model 1 is 71% higher than that of Model 3 and about 1.5% higher than that of Model 2. This means that Model 1 is able to maintain a higher mean driver profit while minimizing the driver profit disparity as compared to Model 3.
 - Fig 4.9 shows that taxis that are serving a recommendation have a higher chance of getting scheduled when the system uses Model 1 than that of Model 2. The scheduling guarantee ratio (SGR) for taxis on average when using Model 1 is 39.8% more than Model 2.

4.5 Conclusion

In this chapter, we first explained the notion of service fairness in the taxi fleet management. We emphasized on two key components of the service fairness, namely the driver profit disparity and the scheduling guarantee. Our objective, with regards to service fairness was to build a model that can firstly, reduce the disparity among the driver earnings while ensuring that the mean driver profits were increased and secondly, increase the booking rate for the drivers that are adhering to the system recommendations. Then we elaborated the proposed multi-objective constrained optimization based taxi fleet management model, to ensure the maximal company profit and the balanced individual drivers' benefits, thus providing a unified framework to ensure taxi scheduling, vehicle rebalancing, and service fairness. In the formulation of this model, several practical constraints were captured by a genuine mathematical setup, which clearly indicates the feasibility of our goal in practice. To curb the computational complexity, we introduced a receding horizon strategy, which allowed us to handle a relatively short decision horizon at each time in a repetitive manner. To fully demonstrate the impact of each objective to the overall system performance and service fairness, we investigated three different scenarios, each of which is equipped with one cost function that captured only some operational goals. The striking simulation case study outcomes clearly showed the usefulness of our taxi fleet management strategy. Using the proposed model, we were able to increase the company level benefits such as the served demands, company profits, etc, and at the same time improve the driver level benefits such as the driver profits and booking rates.

Chapter 5

Summary and Possible Future Directions

5.1 Summary

In the literature, we find numerous references that fleet coordination is a promising upgrade to the common industrial practices in taxi fleet management. The primary reason is that fleet coordination provides a better control of the supply distribution over the operating area. Some of the key challenges in incorporating the fleet coordination are efficient real time implementations of the vehicle rebalancing policy and ensuring service fairness for the drivers while they are working in a coordinated environment. In this thesis, we have addressed these challenges through the proposed Coordinated Mobility-on-Demand fleet management model. First, we discussed the key features of the proposed model, namely the supply scheduling, vehicle rebalancing and CMoD fleet control. The proposed linear programming based models for the supply scheduling and vehicle rebalancing were discussed in detail. Supply scheduling model is used to match the idle standing drivers with the passengers. The objective is to maximize the overall profits that can be earned by serving the current demands. Whereas, the vehicle rebalancing model is used to redistribute the idle standing vehicles with the objective of maximizing the expected served demands in the future at a minimum idle mileage travelling cost. A Receding Horizon Con-

trol (RHC) was used to implement the decision making, i.e., supply scheduling and vehicle rebalancing in a sequential manner. In order to test the efficacy of the proposed CMoD model, we tested its supply-demand imbalance mitigation capability by comparing the two policies, namely the waiting policy in which the taxis wait in their respective locations after serving the current demands and the proposed dynamic supply-demand balancing policy where we redistribute the idle standing vehicles to serve the demands in the future.

We identified two model parameters namely the rebalancing horizon window and rebalancing frequency that are critical for the effective real-time implementation of the vehicle rebalancing model. Upper-Confidence-Bound based algorithm was used to learn the optimal set of vehicle rebalancing model parameters. It was found that the proposed dynamic supply-demand balancing policy based model led to a significant improvement in the served demands as compared to the model that employs the waiting policy.

We then extended our CMoD model to incorporate the service fairness factor. The existing challenges to ensure service fairness for the drivers while operating in the coordinated environment are that the drivers have their own objectives of maximizing their profits. So, the drivers shall adhere to the recommendations by the system if and only if they see the clear benefits. In the proposed upgrade to the model, we formulated a multi-objective cost function aiming for an optimal balance between the company profit and service fairness among individual drivers. We quantified the company benefits in terms of the served demands and the profits. The driver benefits were quantified as the scheduling guarantee and the driver profit disparity. It was observed that the proposed model led to a significant improvement in the company as well as the driver benefits thus ensuring an optimal balance between the company or fleet level performance as well as the individual driver level benefits.

5.2 Future Directions

Some of the possible future directions of the existing works are as follows:

- In order to test the scalability and transfer-ability of the proposed model, the supply-demand balancing model will be tested using data sets from various sources and cities. Apart from that the aspect of robustness to the malicious attacks and forecasting inaccuracies will be pursued in the future works.
- The primary advantage of using the proposed coordinated mobility-on-demand system model is that we can control the supply distribution over the operating area using the vehicle rebalancing strategy. The objective of this strategy is to reduce the supply-demand imbalance. A natural extension to this methodology is to incorporate the strategies to control the demand using dynamic pricing. So, the simultaneous usage of vehicle rebalancing and dynamic pricing can be an improvement to the existing practices in the industry for taxi fleet management.
- The efficacy of supply-demand balancing models depend on the human response to these control signals (recommendations, change in the prices etc). So, another possible research direction is to learn a model of the human response to such signals. If learned correctly this information can be generalized for similar markets and can help in improving the quality of service provided by the MoD systems.

Publications

- A. S. Brar and R. Su, "Ensuring Service Fairness in Taxi Fleet Management," 2020 IEEE 23rd International Conference on Intelligent Transportation Systems (ITSC), 2020, pp. 1-6, doi: 10.1109/ITSC45102.2020.9294504
- A. S. Brar and R. Su, "Dynamic Supply-Demand Balancing Policy for CMoD Fleet," 2021 IEEE 24th International Conference on Intelligent Transportation Systems (ITSC), 2021 - accepted

Bibliography

- [1] F. Miao et al., "Data-Driven Robust Taxi Dispatch Under Demand Uncertainties," in *IEEE Transactions on Control Systems Technology*, vol. 27, no. 1, pp. 175-191, Jan. 2019, doi: 10.1109/TCST.2017.2766042.
- [2] F. Miao et al., "Taxi Dispatch With Real-Time Sensing Data in Metropolitan Areas: A Receding Horizon Control Approach," in *IEEE Transactions on Automation Science and Engineering*, vol. 13, no. 2, pp. 463-478, April 2016, doi: 10.1109/TASE.2016.2529580.
- [3] Pavone M, Smith SL, Frazzoli E, Rus D. Robotic load balancing for mobility-on-demand systems. *The International Journal of Robotics Research*. 2012;31(7):839-854. doi:10.1177/0278364912444766
- [4] A. S. Brar and R. Su, "Ensuring Service Fairness in Taxi Fleet Management," 2020 IEEE 23rd International Conference on Intelligent Transportation Systems (ITSC), Rhodes, Greece, 2020, pp. 1-6, doi: 10.1109/ITSC45102.2020.9294504.
- [5] A. Wallar, M. Van Der Zee, J. Alonso-Mora and D. Rus, "Vehicle Rebalancing for Mobility-on-Demand Systems with Ride-Sharing," 2018 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), Madrid, Spain, 2018, pp. 4539-4546, doi: 10.1109/IROS.2018.8593743.
- [6] J. Alonso-Mora, S. Samaranayake, A. Wallar, E. Frazzoli, and D. Rus, "On-demand high-capacity ride-sharing via dynamic trip-vehicle assignment," *Proceedings of the National Academy of Sciences*, p.201611675, 2017.
- [7] Juan Camilo Castillo, Dan Knoepfle, and Glen Weyl. 2017. Surge Pricing Solves the Wild Goose Chase. *ACM Transactions on Economics and Computation* 1, 1, Article 1 (January 2017), 2 pages. DOI:<http://dx.doi.org/10.1145/3033274.3085098>
- [8] <https://www1.nyc.gov/site/tlc/about/tlc-trip-record-data.page>
- [9] H. Rong et al., "Mining Efficient Taxi Operation Strategies From Large Scale Geo-Location Data," in *IEEE Access*, vol. 5, pp. 25623-25634, 2017.

- [10] Jing Yuan, Yu Zheng, Liuhang Zhang, Xing Xie, and Guangzhong Sun. 2011. Where to find my next passenger. In Proceedings of the 13th international conference on Ubiquitous computing (UbiComp '11). Association for Computing Machinery, New York, NY, USA, 109–118. DOI:<https://doi.org/10.1145/2030112.2030128>
- [11] J. Wen, J. Zhao and P. Jaillet, "Rebalancing shared mobility-on-demand systems: A reinforcement learning approach," 2017 IEEE 20th International Conference on Intelligent Transportation Systems (ITSC), 2017, pp. 220-225, doi: 10.1109/ITSC.2017.8317908.
- [12] Jiao Y, Tang X, Qin ZT, Li S, Zhang F, Zhu H, Ye J. A Deep Value-based Policy Search Approach for Real-world Vehicle Repositioning on Mobility-on-Demand Platforms.
- [13] Yuan NJ, Zheng Y, Zhang L, Xie X. T-finder: A recommender system for finding passengers and vacant taxis. IEEE Transactions on knowledge and data engineering. 2012 Aug 6;25(10):2390-403.
- [14] Qu M, Zhu H, Liu J, Liu G, Xiong H. A cost-effective recommender system for taxi drivers. In Proceedings of the 20th ACM SIGKDD international conference on Knowledge discovery and data mining 2014 Aug 24 (pp. 45-54).
- [15] A. Sadeghi and S. L. Smith, "On Re-Balancing Self-Interested Agents in Ride-Sourcing Transportation Networks," 2019 IEEE 58th Conference on Decision and Control (CDC), 2019, pp. 5119-5125, doi: 10.1109/CDC40024.2019.9030043.
- [16] Wollenstein-Betech S, Paschalidis IC, Cassandras CG. Joint pricing and rebalancing of autonomous mobility-on-demand systems. In 2020 59th IEEE Conference on Decision and Control (CDC) 2020 Dec 14 (pp. 2573-2578). IEEE.
- [17] Ruch C, Hörl S, Hakenberg J, Frazzoli E. The impact of fleet coordination on taxi operations.
- [18] Hörl S, Ruch C, Becker F, Frazzoli E, Axhausen KW. Fleet operational policies for automated mobility: A simulation assessment for Zurich. Transportation Research Part C: Emerging Technologies. 2019 May 1;102:20-31.
- [19] J. Xu, R. Rahmatizadeh, L. Bononi and D. Turgut, "Real-Time Prediction of Taxi Demand Using Recurrent Neural Networks," in IEEE Transactions on Intelligent Transportation Systems, vol. 19, no. 8, pp.2572-2581, Aug. 2018.[6] L.
- [20] Moreira-Matias, J. Gama, M. Ferreira, J. Mendes-Moreira and L.Damas, "Predicting Taxi-Passenger Demand Using Streaming Data," in IEEE Transactions on Intelligent Transportation Systems, vol. 14, no. 3, pp. 1393-1402, Sept. 2013.

- [21] <https://www.nissan.com.sg/vehicles/new/nv200/specifications.html>
- [22] https://www.eia.gov/dnav/pet/pet_pri_gnd_dcus_y35ny_w.htm
- [23] C. Ruch, J. Gächter, J. Hakenberg and E. Frazzoli, "The +1 Method: Model-Free Adaptive Repositioning Policies for Robotic Multi-Agent Systems," in *IEEE Transactions on Network Science and Engineering*, vol. 7, no. 4, pp. 3171-3184, 1 Oct.-Dec. 2020, doi: 10.1109/TNSE.2020.3017526.
- [24] Jiaqi Ma, Xiaopeng Li, Fang Zhou, Wei Hao, "Designing optimal autonomous vehicle sharing and reservation systems: A linear programming approach", *Transportation Research Part C: Emerging Technologies*, Volume 84, 2017, Pages 124-141, ISSN 0968-090X, <https://doi.org/10.1016/j.trc.2017.08.022>.
- [25] R. Zhang, K. Spieser, E. Frazzoli and M. Pavone, "Models, algorithms, and evaluation for autonomous mobility-on-demand systems," 2015 American Control Conference (ACC), 2015, pp. 2573-2587, doi: 10.1109/ACC.2015.7171122.
- [26] Maciejewski, M.; Nagel, K. (2014). The Influence of Multi-agent Cooperation on the Efficiency of Taxi Dispatching. *PPAM 2013: Parallel Processing and Applied Mathematics*, 751–760.https://doi.org/10.1007/978-3-642-55195-6_71
- [27] R. Zhang, F. Rossi and M. Pavone, "Model predictive control of autonomous mobility-on-demand systems," 2016 IEEE International Conference on Robotics and Automation (ICRA), 2016, pp. 1382-1389, doi: 10.1109/ICRA.2016.7487272.
- [28] Hamid R. Sayarshad, Joseph Y.J. Chow, Non-myopic relocation of idle mobility-on-demand vehicles as a dynamic location-allocation-queueing problem, *Transportation Research Part E: Logistics and Transportation Review*, Volume 106, 2017, Pages 60-77, ISSN 1366-5545, <https://doi.org/10.1016/j.tre.2017.08.003>.
- [29] Chen, M. Keith, and Michael Sheldon. "Dynamic Pricing in a Labor Market: Surge Pricing and Flexible Work on the Uber Platform." *Econometrica* 84.5 (2016): 2940-2978.
- [30] Manski, Charles F., and J. David Wright. Nature of equilibrium in the market for taxi services. No. Intrm Rpt.. 1967.
- [31] Frechette, Guillaume R., Alessandro Lizzeri, and Tobias Salz. "Frictions in a competitive, regulated market: Evidence from taxis." *American Economic Review* 109.8 (2019): 2954-92.
- [32] Ge, Yong, et al. "An energy-efficient mobile recommender system." *Proceedings of the 16th ACM SIGKDD international conference on Knowledge discovery and data mining*. 2010.

- [33] Rong, Huigui, et al. "Mining efficient taxi operation strategies from large scale geo-location data." *IEEE Access* 5 (2017): 25623-25634.