

ABSTRACT

Title of Thesis: AN INTEGER PROGRAMMING MODEL
FOR DYNAMIC TAXI-SHARING
CONSIDERING PROVIDER PROFIT

Yeming Hao, Master of Science, 2018

Thesis Directed By: Professor Ali Haghani
Department of Civil and Environmental
Engineering

This thesis proposes an integer programming model for Dynamic Taxi-Sharing (DTS), which allows two groups of taxi users to ride on the same taxi together. The model matches taxi drivers and user pairs in certain sequences with the goal of maximizing taxi providers' profit. We also develop a DTS fare calculation scheme which can automatically calculate the fare for each DTS user and self-adjust to balance the taxi occupancy rate in real time. A customized spectral clustering approach for preselection on DTS trips is also designed to narrow down the search space for the model. Real-world taxi trip data is used to demonstrate the DTS system is beneficial to providers, taxi users, and taxi drivers.

AN INTEGER PROGRAMMING MODEL FOR DYNAMIC TAXI-SHARING

CONSIDERING PROVIDER PROFIT

By

Yeming Hao

Thesis submitted to the Faculty of the Graduate School of the
University of Maryland, College Park, in partial fulfillment
of the requirements for the degree of
Master of Science
2018

Advisory Committee:

Professor Ali Haghani, Chair

Professor Paul Schonfeld

Dr. Kaveh Farokhi Sadabadi

© Copyright by
Yeming Hao
2018

Acknowledgments

I would first like to thank my academic advisor Professor Ali Haghani sincerely. I benefited a lot from his insights, guidance, support, patience, and valuable advice. He is not only a mentor in research but also a mentor in life. He welcomed me into his research group while I was just fresh out of undergrad, knowing very limited knowledge in the field. Seeing him patiently guiding every student and witnessing his calmness in all unexpected circumstances over these years deeply impressed me. I feel honored to have the opportunity to conduct my study under his supervision and learn the wisdom of life from him.

I also need to thank all the professors/course instructors from whom I had the honor to learn knowledge in and out of class: Professor Paul Schonfeld, who is a member of my advisory committee and gave me constructive comments, Professor Lei Zhang, Professor Cinzia Cirillo, Professor Elise Miller-Hooks, Professor Bruce Golden, Professor Frank B. Alt, Dr. Charles C. Liu, Dr. Mohamadreza Banihashemi, Professor Rama Chellappa, Professor Ilya Ryzhov, and Professor Mark Austin.

I want to express my special thanks to Professor Lei Zhang, the Director of the National Transportation Center at UMD and Professor Paul Schonfeld, the faculty representative at the Intelligent Transportation Society of America and Institute of

Transportation Engineers University of Maryland Student Chapter (ITS-ITE UMD Student Chapter). They supported me while I was working as an event coordinator and then the president of the student chapter, representing the chapter and exploring new cooperation with different academic and industry organizations/professionals. Those valuable experiences helped me get to know the field better and step out of campus with real-world connections to better serve the faculty/student community here in our department.

My sincere thanks also go to Dr. Kaveh Farokhi Sadabadi, who is also a member of my advisory committee and other colleagues in CATT (Center for Advanced Transportation Technology) including Mr. Thomas Jacobs, Mr. Zach Vander Laan, Dr. Sepideh Eshragh, Dr. Golnush Masghati Amoli, , Dr. Nikola Markovic, Dr. Elham Sharifi, Dr. Masoud Hamed and others, for the trust and the opportunity to work with them as a graduate research assistant and participate in many exciting real-world projects. Dr. Kaveh Farokhi Sadabadi selflessly guided me through work, study, and life these years. I feel lucky to have such a dedicated and brilliant mentor from whom I could learn.

I also owe a debt of gratitude to my family: my parents, my elder sister, and my twin sister Shuming Hao, for supporting me spiritually throughout writing this thesis and my life in general. Shuming and I chose to pursue graduate degrees at the same time,

although in two completely different fields. Her dedication and intelligence are truly an inspiration for me. I feel truly grateful that she is always there encouraging and supporting me.

Last but not least, I would like to thank all my friends, lab mates, workout partners, especially Zhongxiang Wang, Sachraa Gangsa, Dr. Liang Liang, Yuntao Liu, Dr. Xuechi Zhang, Binya Zhang, Qian Wang, Guowei Sun, Haoshu Duan, Dr. Stefan Poikonen, Wenmo Sun, Moschoula Pternea, Kiana Roshan Zamir, Dr. Rafael Olarte, Sanaz Aliari, Dr. Bahar Zarin, Luanjiao Hu and many others for the time we spent together, all the stimulating discussions and all the priceless support.

Table of Contents

Acknowledgments.....	ii
Table of Contents	v
List of Tables	ix
List of Figures	xi
Chapter 1: Introduction.....	1
1.1 Research Background	1
1.2 Advantages.....	4
1.3 DTS System Structure.....	6
1.4 Contribution	10
1.5 Structure of Thesis	11
Chapter 2: Literature Review.....	12

2.1 Classifications	12
2.1.1 Traditional Ridesharing	12
2.1.2 Dynamic Ridesharing.....	14
2.1.3 Dynamic Taxi-Sharing.....	23
2.1.4 Summary of the Three Ridesharing Services.....	29
2.2 Model Features.....	31
2.2.1 Ridesharing Objectives	31
2.2.2 Matching Constraints.....	32
2.2.3 Dynamic Strategy.....	33
2.2.4 Cost Sharing/Fare Splitting Scheme	34
2.3 Chapter Conclusion.....	38
Chapter 3: Model Formulation.....	40
3.1 Notation.....	42
3.2 Objective Function.....	43

3.3 Constraints	50
3.4 Chapter Conclusion.....	58
Chapter 4: DTS Fare Calculation Scheme	59
4.1 The Basics	59
4.2 The Dynamics	61
4.3 Example	62
4.4 Chapter Conclusion.....	67
Chapter 5: A Customized Spectral Clustering Approach for Preselection on DTS Trips	68
5.1 Background	68
5.2 Approach Design	73
5.3 Chapter Conclusion.....	79
Chapter 6: Case Study.....	80
6.1 Data Description and Preparation	80

6.2 Implementation Setup	85
6.2.1 Approach and Environment	85
6.2.2 Parameters	86
6.3 Implementation Results	92
6.3.1 Preliminary Experiment	92
6.3.2 Case Study I	92
6.3.3 Case Study II	100
6.3.4 Case Study III	105
6.4 Chapter Conclusion	106
Chapter 7: Summary, Conclusions, and Future Research	108
7.1 Summary and Conclusions	108
7.2 Future Research	110
References	113

List of Tables

Table 1 Solution Values for a Large Amount of Carpooling Inquiries	23
Table 2 Summary of the Features of Three Ridesharing Services.....	29
Table 3 Data Sets for DTS	42
Table 4 Decision Variables for DTS.....	42
Table 5 Parameters for DTS	42
Table 6 Fare Comparison before and after DTS for the Two Users.....	66
Table 7 Fare Comparison before and after DTS for the Driver.....	67
Table 8 Sample of the TLC Trip Record Data Set (Part 1/2)	81
Table 9 Sample of the TLC Trip Record Data Set (Part 2/2)	82
Table 10 Data Dictionary.....	82
Table 11 Trip Distance Ratio Calculation Sample	89
Table 12 Parameters Values.....	90

Table 13 Results for Different ρ Values in Time Interval 1	94
Table 14 Results for Time Interval 2	98
Table 15 Results for Time Interval 3	99
Table 16 Results for Time Interval 1	101
Table 17 Results for Time Interval 2	102
Table 18 Results for Time Interval 3	104
Table 19 Results with and without 2 nd Level Clustering	105

List of Figures

Figure 1 Percent of Shifts Started by Time of Day (15-minute Increments).....	2
Figure 2 Average Percentage of Taxis Occupied by Time of Day (15-minute Increments)	3
Figure 3 DTS System Structure	7
Figure 4 Qualitative Results of the Satisfaction Survey for Taxi-sharing among Government, Operators, Drivers, and Passengers	25
Figure 5 Dividing of the Simulation Area	26
Figure 6 Data Set	44
Figure 7 Drop-off Sequence.....	48
Figure 8 Balance System-wide Taxi Occupancy Rate.....	62
Figure 9 User No. 30 Route from Google Maps.....	63
Figure 10 User No. 16 Route from Google Maps.....	64
Figure 11 the DTS Trip Route	65

Figure 12 Midpoint	69
Figure 13 Azimuth in Navigation	70
Figure 14 Normalized Vector Pairs	74
Figure 15 the Function of $\cos x$	74
Figure 16 Flowchart of the Approach in One Time Interval	78
Figure 17 Boro Taxis and Yellow Taxis Cover Area	81
Figure 18 Location of the Pick-up Points	84
Figure 19 Location of the Drop-off Points	85
Figure 20 Using Haversine Formula to Calculate the Great-circle Distance.....	86
Figure 21: A Rectangular Plane Intersecting a Great Circle Path and the Center of the Earth (Arc b is the Path and the Angular Separation of the End Points).....	87
Figure 22 Calculation Time for Different Data Set Sizes.....	92
Figure 23 Clustering Results for Time Interval 1	94
Figure 24 Total Profit for Different ρ Values.....	96

Figure 25 Total Driver Benefit Change for Different ρ Values	96
Figure 26 Different Rate Change for Different ρ Values	97
Figure 27 Clustering Results for Time Interval 2	98
Figure 28 Clustering Results for Time Interval 3	99
Figure 29 Clustering Results for Time Interval 1	101
Figure 30 Clustering Results for Time Interval 2	102
Figure 31 Clustering Results for Time Interval 3	104

Chapter 1: Introduction

1.1 Research Background

Taxi transportation has become an indispensable part of today's public transportation system. Similar to private vehicles, taxis have high mobility and high accessibility. However, taxi transportation also raises public's attention by bringing more traffic congestion, fuel consumption, air pollution and high personal travel cost. Solving these problems is crucial to the future development of taxi transportation.

According to New York 2014 Taxicab Fact Book, there are on average 485,000 yellow taxi trips, and 600,000 passengers per day (Bloomberg & Yassky), which gives us an average of 1.24 passengers per taxi trip. Assuming a typical taxi has four passenger seats, this leaves 2.76 unoccupied passenger seats per taxi trip and 1,338,600 unoccupied seats for all the yellow taxi trips in New York in one day, which means there are 489 million unoccupied seats for all the yellow taxi trips in New York in one year. Effectively using these unoccupied seats is a vital point in the current transportation situation.

Meanwhile, taxi riders have difficulty getting taxi cabs during peak hours in many major cities. There is new data to confirm what generations of New Yorkers have long known in their bones: just as the afternoon rush is about to begin, the taxicabs disappear by the hundreds. From 4 to 5 p.m., the traditional hour for taxicabs to

change shifts, the number of active taxicabs on the streets falls by nearly 20 percent compared with an hour before (Grynbaum, 2011).

2014 Taxicab Fact Book(Bloomberg & Yassky) also gives us the shift information with real-world taxi data as follows:

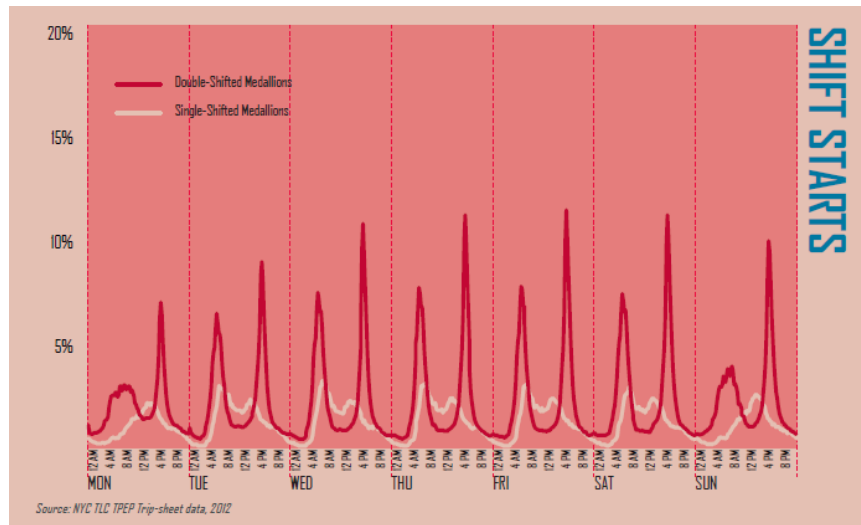


Figure 1 Percent of Shifts Started by Time of Day (15-minute Increments)

As we can see from Figure 1, fleet vehicles tend to start their shifts around a centralized time for both the AM and PM shifts. This is especially true on weekdays when on average, 39% of vehicles operated under this model start their evening shifts in the 5:00-6: 00 PM hour, with over 10% starting in just the 5:00-5:15 block alone. The morning shift start times are also clustered to a degree, but less than the evening shift. In the morning shift, 28% of fleet vehicles begin weekday AM shifts in the 6:30-7: 30 PM hour, on average (Bloomberg & Yassky).

As shown in Figure 2, there is a daily spike in the percentage of available taxis that are occupied between 4 PM and 6 PM each day. On average, 64% of taxis are occupied during these hours (Bloomberg & Yassky).

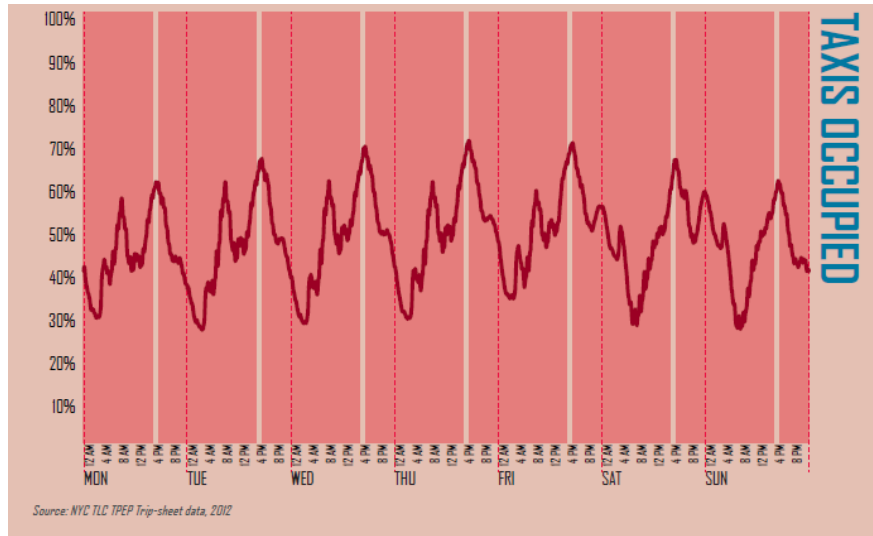


Figure 2 Average Percentage of Taxis Occupied by Time of Day (15-minute Increments)

Increasing the number of taxicabs seems like an obvious solution to the lack of taxicabs during peak hours. However, this will cause more traffic congestion for the entire road network. Considering the uneven distribution of taxi occupancy rate and the large number of unoccupied seats, a dynamic taxi-sharing (DTS) service which can allow more than one taxi users ride on the same taxi together and can be adjusted to balance the real-time taxi occupancy rate is necessary to alleviate the issue.

Technological development is the feasibility enabler to the DTS idea. 64% of American adults now own a smartphone, up from 35% in 2011 (Smith, 2015). Two-

thirds of New York taxi passengers own or use a smartphone. 55% say they would like the option of using their phone to locate taxicabs, and 54% say they would pay for their rides with their phone if they could do so (Bloomberg & Yassky). The growing ubiquity of Internet-enabled mobile devices partially enables practical dynamic ridesharing (Hartwig & Buchmann, 2007).

Although the technology is already available, DTS has not been well studied and applied. The development of algorithmic approaches for optimally matching taxi drivers and users in real-time and the corresponding taxi fare calculation scheme is very central to the development of the concept.

1.2 Advantages

The DTS can bring many advantages to the taxi providers, users, and the society as follows:

(1) The decrease of personal travel cost

Sharing the taxi ride with others allows riders to save money on the taxi fare.

(2) The increase in users

The price elasticity of demand for personal transportation services is very high. When Uber (an online transportation network company which allows consumers with smartphones to submit a trip request which is then routed to Uber drivers who use

their own cars) launched its low-cost UberX offering in the summer of 2012, the company quickly realized that the demand for its transportation services is highly elastic. As the company achieved lower and lower per-ride price points, the demand for rides increased dramatically. A lower price point delivered a much better value proposition to the consumer, yet remained a great business decision due to the remarkable increase in demand (Deamicis, 2015).

Real data has already proved that similar products to DTS can help the taxi service providers gain more users from the launching of a DTS similar product – UberPOOL (a carpool service Uber offers which will be further introduced in Chapter 2).

According to the data Uber released on its blog, until April 2015, millions of trips have been taken on UberPOOL since it launched in August 2014. Thousands take it five times a week during commuting hours in the cities where UberPOOL is active. In some concentrated neighborhoods, match rates during this time of day are at 90 percent (Myhrvold, 2015).

(3) Fewer cars on the road and fewer CO₂ emissions

Dynamic Taxi-Sharing could be part of the solution for urban transportation congestions and car emissions. The taxi providers will need fewer drivers to serve the same amount of taxi users and thus reduce the emission and help elevate the congestion during peak hours. Take UberPOOL data as an example, if we assume the UberPOOL riders' alternative method is individual personal Uber rides, the miles

savings estimate for San Francisco (the distance difference between the sum of the individual rider routes and the UberPOOL route) is about 674,000 miles from February 20th to March 20th, 2015. Conservatively assuming that every San Francisco UberX vehicle is a Toyota Prius with the gas mileage of 50 mpg, UberPOOL trips saved around 13,500 gallons of gasoline. Accounting for a savings of 8.91 kg of atmospheric CO₂ emissions per gallon, San Francisco UberPOOL prevented about 120 metric tons of CO₂ emissions from February 20th to March 20th, equivalent to the output of over 128,000 pounds of coal (Myhrvold, 2015).

(4) More social opportunities for users

DTS service allows more users in a taxi and thus can give taxi riders and drivers more opportunities to talk to each other. The service sometimes even functions like a blind-date and opens the door to romance. Taking a similar product UberPOOL as an example; Uber says that at least one couple, who they identify only as Oliver and Jennifer, are now engaged after they found themselves riding together to the same restaurant in San Francisco (Wagstaff, 2015).

1.3 DTS System Structure

We use the term *Dynamic Taxi-Sharing (DTS)* to describe an automated system that facilitates taxi users to share one-time taxi trips in their desired travel times with other taxi riders heading in a similar direction. The system automatically matches riders

and drivers with certain sequences and calculates the fare for the shared taxi trip. The system will also adjust its parameters to balance the real-time taxi occupancy rate in the area.

We assume the DTS system (shown in Figure 3) consists of three parts which can be connected to and communicate with each other through the internet:

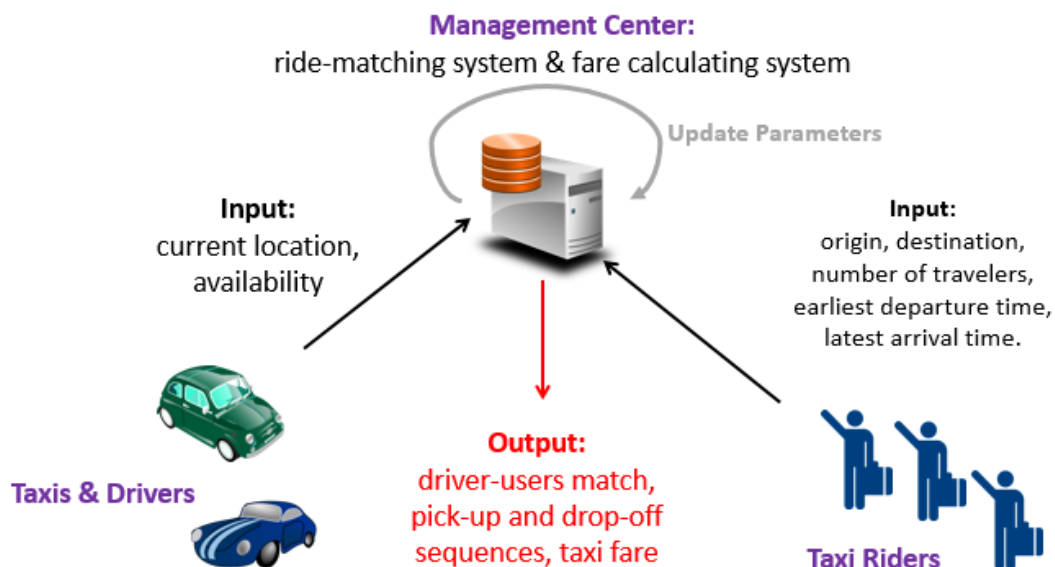


Figure 3 DTS System Structure

(1) Taxis and taxi drivers provide current location and availability. The locations of all taxis in the system are tracked and monitored by GPS. The On-board Units (can be mobile phones) automatically gather this information at a certain time window and send updates to the management server. The drivers will

receive the matched taxi users' information (pick-up and drop-off location, preferred time window, contact information) after being matched.

- (2) Taxi users order DTS service using their mobile phones. They enter their travel origins, destinations, number of travelers (limited to two at most), earliest and latest arrival time through a mobile app. The users will receive the matched taxi driver's information (current location, estimated arrival time, vehicle information, contact information) after being matched. They have the final option to accept or reject the match.
- (3) Management center operates a server that receives all the information from taxi drivers and taxi users, matches users to taxi drivers, calculates the estimated fare for each taxi user and sends all the information back to the drivers and users in real time. If both taxi driver and users accept a proposed match, the driver will pick up and drop off the taxi users in a certain assigned sequence. The server will also automatically collect the fare from all users to the taxi drivers. The location, status of the drivers and users, will also be monitored during the DTS ride and sent back to the server so that the taxi provider can track and confirm the service status at all times. The occupancy of all the taxis in the system will also be monitored.

There are many challenges to implementing this DTS system. The key point of the DTS system is how to optimally match the taxi drivers and users on behalf of the taxi providers and automatically calculate the taxi fare for each user.

1.4 Contribution

This thesis develops a model that gives the matching results for DTS taxi drivers and users; each matching result will contain both pick-up and drop-off sequences, as well as develop the corresponding fare calculation scheme. The main contributions of this thesis can be summarized as follows:

- We propose a matching model to match taxi drivers and user pairs on behalf of taxi providers. The taxi providers can maximize their profit by launching the DTS system.
- We develop a taxi fare calculation scheme which gives both taxi drivers and riders monetary incentive to use the DTS system. Besides, the scheme can self-adjust according to the current taxi occupancy rate to balance the occupancy rate over time.
- We introduce a clustering approach to narrow down the search space for the integer programming model so that the model can give matching results more efficiently.
- We perform case studies to validate the feasibility and the effectiveness of the DTS system.

1.5 Structure of Thesis

The organization of this thesis is as follows. In Chapter 2, we summarize the previous study in this field. In Chapter 3, we develop the integer programming model to match the drivers and users optimally. In Chapter 4, we design the scheme to calculate the DTS fare for each participant. In Chapter 5, we propose a customized structural clustering approach for Preselection on DTS Trips. In Chapter 6, we perform case studies to test the model and the clustering approach. In Chapter 7, we provide our main summary, conclusions and discuss the future research.

Chapter 2: Literature Review

The study on dynamic taxi sharing is still limited. We believe DTS has some similar features to traditional ridesharing and dynamic ridesharing which has been studied for years. This chapter aims to review and analyze the existing studies focused on all the three ridesharing modes.

2.1 Classifications

2.1.1 Traditional Ridesharing

The first use of carpool matching assistance in the United States of America occurred during World War II. The Federal Office of Defense Transportation noted that the average number of passengers per vehicle stood at less than two in 1942. Ridesharing was promoted in response to gasoline and tire rationing (U.S. Office of Defense Transportation, 1942).

One carpooling method sponsored by the U.S. Office of Civilian Defense was called the "Car Sharing Club Exchange and Self-Dispatching System." Participants would fill in a card at an exchange office that included information such as address, commute hours, and phone number along with whether seeking a ride or passengers. A member of the transportation committee operating the exchange would then place the card in the appropriate zone that represented the area in which the person lived.

Riders were free to look through the cards to find rides or passengers that best suited their needs. All ridesharing arrangements and possible fees were made between the individuals, not the exchange office. Of course, officials encouraged participants to be deliberate about spelling out the responsibilities of each member of the carpool.

Issues such as lateness, bad manners, or personal hygiene deficiencies could lead to friction, or worse (U.S. Office of Civilian Defense, 1942).

Carpooling and vanpooling were also promoted extensively during the energy crises and the oil embargo in the mid-1970s and early 1980s. Many employers and regional agencies initiated rideshare programs and carpool matching services during this time, and federal funding was used to support many of these efforts (Turnbull, 2000).

Similar to how financial institutions operated at that time, traditional ridesharing was arranged on bulletin boards at local matching institutions. Users accessed these systems by providing the necessary information over the telephone or by mailing in an application form. They were then offered the potential ridesharing partners' information so they could contact each other and form the ridesharing themselves. This was inconvenient for users and thus affected the successful match rate. Also, traditional ridesharing was not able to handle real-time and flexible needs.

2.1.2 Dynamic Ridesharing

As new technologies such as mobile technology and global positioning system (GPS) evolved, studies considering ridesharing in a dynamic, real-time setting appeared.

Dynamic ridesharing is defined as two or more people sharing a single, non-recurring trip, without regard to previous arrangements or history among the individuals involved and do not require long-term commitments. The trips are prearranged (but on short notice) which means that the participants agree to share a ride in advance, typically while they are not yet at the same location (Agatz, Erera, Savelsbergh, & Wang, 2012). In comparison to traditional ridesharing services, which focus on commuters traveling to and from the same origins and destinations on fixed schedules, a dynamic ridesharing system must be able to match random trip requests at any time (Dailey, Loseff, & Meyers, 1997; Turnbull, 2000).

The real-time or instant carpool concept was tested initially in the Seattle area as part of the Bellevue Smart Traveler (BST) project (Pieratti, Haselkorn, & Blumenthal, 1993). The goal of BST was to design and test an information system that would help decrease single-occupancy vehicles (SOV) travel to a downtown employment center by making alternative commuting options more attractive and easier to access.

The BST Traveler Information Center (TIC) integrated phone and paging technology to deliver three types of personal commuter information: (1) dynamic ride matching information, (2) up-to-the-minute traffic congestion information, and (3) transit

information. Registered users with pagers could view a list of riders offered and current traffic reports while guest users could only access the system using the telephone.

The demonstration period lasted from late November 1993 to late April 1994. At the program's peak, 53 users were registered. 48 of them formed three ride groups. Members of the ride groups offered 509 rides. By telephone, the 48 ride group members looked for 148 rides and accessed additional information on 33 specific rides. However, only six ride matches were logged. (Note that logging a ride was optional, so that ride matches could have occurred without being logged.) Participants liked the idea but were either unable or unwilling to form ride matches. And far more people were interested in inviting others into their car than they were in getting into someone else's car (Pieratti et al., 1993).

We believe the main reason for people's unwillingness to participate in the rideshare is they do not have enough incentives. Creating incentives such as cost-sharing, and high-occupancy vehicle lane (HOV) use for ridesharing is necessary for future attempts. Adjusting the technology accordingly is also essential. In the Seattle experiment, although the idea "dynamic" was used, the experiment wasn't able to follow up on people's real-world dynamic needs using just telephones and pagers. Web-based technologies that can be updated more timely and record these ridesharing trips will be helpful in the ridesharing industry. Moreover, web-based technologies

can help maintain user profiles and trip histories so people will have fewer safety concerns about getting into others' cars.

The Seattle Smart Traveler (SST) tested a dynamic ride-matching system using the Internet and electronic mail (email) at the University of Washington in Seattle from 1995 to 1997. The SST was designed to meet the needs of individuals interested in forming ongoing carpooling arrangements, as well as those interested in offering or obtaining a ride for a single trip. Participants completed SST application forms on the website. Three types of potential matches could be requested. These were regular commute trips, additional regular trips, and occasional trips. A user entered the origin, destination, day of the week, departure time window, and arrival time window for each trip. A search structure was developed allowing users to identify their desired origins and destinations from a search tree containing four levels of details. The system then identified potential matches. The SST automatically generated and sent an email message with this information if the user desired or the participant could call the potential matches. Implementation and operation of the SST lasted for a 15-month period from mid-March 1996 to June 1997. Approximately 400 individuals registered for the SST, with 200 as the largest number of active participants during the peak period. The SST database was updated at the start of each quarter (Dailey et al., 1997; Turnbull, 2000).

SST was a successful attempt in combining both traditional ridesharing and dynamic ridesharing since the user group was relatively large at that time, but like more traditional ridesharing services, making the actual connection with potential ridesharing partners was left up to the SST participants. The automatic email feature enhanced the ease of communicating with possible matches but did not alleviate the need for participants to take action themselves (Dailey et al., 1997; Turnbull, 2000).

As the Internet developed, web-based dynamic ridesharing studies appeared. Dobrosielski et al. (2007) built a website for University of Maryland, College Park commuters with the main goal of simplifying group travel for purposes of safety, resource efficiency, and flexibility. The study is mostly a website interface design. Users can search for a carpool or create a new one. Carpool matches are shown visually on the map. The matches are also listed with their ranking criteria, to the right of the map. There are three methods that users can select to rank the matches, and each selection implicitly reorganizes the results. The first is "Time Deviation," which is the deviation from the user's desired arrival time. The second is "Time Added to Carpool," which is the time that is added to the carpool if the user's destination is added to the current list of stops. The third is "Your Total Travel Time," which is the total time the carpool could take from the user's stop to the user's destination. Users may choose any carpool to view its details or to request to join it. This webpage serves more like a platform to gather and show carpool information

and let users choose carpool partners themselves without achieving a systematic goal for the entire carpool community (Dobrosielski, Gray, Nhan, & Stolen, 2007).

Geisberger et al. (2010) designed a pruning strategy to match ridesharing requests and offers. The road network was modeled as a weighted graph, and the edge weights were travel time between the nodes. Results showed that the algorithm is perfectly suitable for a large scale web service with potentially hundreds of thousands of users each day (Geisberger et al., 2010).

Agatz et al. (2011) developed optimization-based approaches that aimed at minimizing the total system-wide vehicle miles incurred by system users and their individual travel costs. A simulation study was also implemented based on the 2008 travel demand data from metropolitan Atlanta. Results indicated that the use of sophisticated optimization methods instead of simple greedy matching rules substantially improves the performance of ridesharing systems. Dynamic ridesharing may have the potential for success in large US metropolitan areas, with sustainable ride-share populations forming over time even with relatively small overall participation rates and when considering only home-based work trips. Besides travel costs savings, ridesharing systems may provide travel time savings to participants by providing access to high occupancy lanes. Moreover, ridesharing may help to decrease traffic congestion and thereby reduce system-wide travel times (N. A. H. Agatz, A. L. Erera, M. W. P. Savelsbergh, & X. Wang, 2011).

Yan et al. (2011) employed a network flow technique to develop a long-term many-to-many carpooling model systematically. The model was formulated as a special integer multiple-commodity network flow problem. A Lagrangian relaxation-based algorithm was also developed to solve the model. The performance of the heuristic algorithm was evaluated by carrying out a case study using real data and suitable assumptions. The test results confirmed the usefulness of the model and the heuristic algorithm and that they could be useful in practice. This study extended the existing “fairness” concept among the participants. Instead of merely considering the frequency of being a driver, they also included the systematic costs (including driver/passenger traveling costs and driver operating costs) for the “fairness” among the participants. This study also extended the pooling of “individuals” to “groups of members that should be assigned to the same car,” which is closer to the situation that occurs in practice. Each participant group would provide a vehicle, and the model would find each participant group’s role (driver group or passenger group), driver group (vehicle) routes, driver group (vehicle) arrival/departure times for all stations, passenger group routes, passenger group boarding and getting-off times, and which passenger group should take which vehicle. The model was formulated to minimize the maximum cost of each person (Yan, Chen, & Lin, 2011).

Kammerdiener & Zhang (2011) described the formulation of algorithms for measuring the closeness of a match between pairs of potential partners in a university

online ridesharing or carpooling system. A measure of “goodness” is defined between two users based upon the difficulty of the trip from one user, via the other user, to the common destination (Kammerdiener & Zhang, 2011).

Ghoseiri (2012) developed a Dynamic Rideshare Optimized Matching (DROM) model and solution that is aimed at identifying suitable matches between passengers requesting rideshare services with appropriate drivers available to carpool for credits and HOV lane privileges. The model was designed to maximize the total number of matching in a given planning horizon while the total passenger and driver travel times are minimized. The research developed a spatial, temporal, and hierarchical decomposition solution strategy that leads to the heuristic solution procedure, Three-Spherical Heuristic Decomposition Model (TSHDM). A case study was constructed to analyze the model and TSHDM behaviors on a road network of the northwest metropolitan area of Baltimore city. Results showed it is possible to implement a dynamic rideshare system using appropriate technology tools and social networking media (Ghoseiri, 2012).

Huang et al. (2015) developed a genetic-based carpool route and matching algorithm (GCRMA) for the multi-objective optimization problem called the carpool service problem (CSP). The paper focused on solving the problem by dramatically acquiring optimal match solutions while reducing the required computing time. Evaluation of the model and algorithm was accomplished by using test scenarios simulating real-world environments. The experimental section showed that the proposed GCRMA

was compared with two single-point methods: the random-assignment hill climbing algorithm and the greedy assignment hill climbing algorithm on real-world scenarios. Use of the GCRMA was proved to result in superior results involving the optimization objectives of CSP than other algorithms. Furthermore, the GCRMA operated with a significantly smaller amount of computational complexity to produce the match results in the reasonable time, and the processing time was further reduced by the termination criteria of the early stop. The primary objective of the CSP was to maximize the total number of passengers matched with drivers, as well as their cumulative credit scores (The social terms and ratings of each user were systematically normalized as a credit score, by which to establish interpersonal trust and responsibility in the carpool system). The secondary objective was to minimize the average travel distance of drivers, the average waiting time of passengers, and the average travel distance of passengers (S. C. Huang, Jiau, & Lin, 2015).

Stiglic et al. (2015) investigated the potential benefits of introducing meeting points in a ridesharing system. Riders could be picked up and dropped off at a meeting point within a certain distance from their origin or destination. The increased flexibility resulted in additional feasible matches between drivers and riders and allowed a driver to be matched with multiple riders without increasing the number of stops the driver needed to make. Maximizing the number of matches and maximizing the driving distance savings were both taken into account in a hierarchical fashion, where

they considered the first one as the primary objective and second one as the secondary objective. Time flexibility of drivers, riders, and departure times were also considered in the study. An extensive simulation study was performed, and the results demonstrated that meeting points could significantly increase the number of matched participants as well as the system-wide driving distance savings in a ridesharing system. (Stiglic, Agatz, Savelsbergh, & Gradisar, 2015).

C. Huang et al. (2016) presented a two-stage integer programming formulation for the carpooling problem. Next, they proposed a stochastic Tabu search (TS) algorithm to solve this problem. The proposed algorithm aimed at a wide range of passenger distribution and routing problems. The computational results based on real-world user data showed the effectiveness of the proposed algorithm. Moreover, they developed a mobile application based on their carpooling model. The objective function was to minimize the total cost and assign the passengers to their nearest driver. In their case study for a large amount of carpooling inquiries, the number of participants (P) was set to a number between 200 and 500. The number of potential drivers (D) was chosen from the range [100, 275]. The maximum allowable number of drivers (K) was selected from the range [50, 137]. Their computation results are reported in Table 1. We can find that the CPU time increases rapidly with the increase in the carpooling participants (C. Huang, Zhang, Si, & Leung, 2016).

Table 1 Solution Values for a Large Amount of Carpooling Inquiries

Instance	P	D	K	ini	TS			
					avg	bst	$t(s)$	$\%gap$
R112	200	100	50	97220.27	26966.08	26481.36	2.0442	72.26
R113	250	125	62	133123.96	35314	34718.85	4.7652	73.47
R114	300	150	75	143870.83	38222.51	37897.99	9.706	73.43
R115	350	175	87	180258.2	46323.44	46032.51	17.1914	74.30
R116	400	200	100	213664.32	53763.75	53258.06	29.5188	74.84
R117	450	225	112	223827.47	56693.82	56216.69	45.2452	74.67
R118	500	250	125	285616.88	69494.29	69165.42	71.3828	75.67
R119	550	275	137	346658.19	82288.76	81833.46	103.693	76.26

This algorithm was effective in the study but may have limitations when applied to real-world situations since the data they used was simulated data, and they did not consider time window in the study.

2.1.3 Dynamic Taxi-Sharing

Agatz et al. (2012) defined dynamic ridesharing as an automated system that facilitates drivers and riders to share one-time trips close to their desired departure times. Dynamic Taxi-Sharing (DTS) contains some features similar to Dynamic Ridesharing but also differs from it because DTS does not require participants have a vehicle and the DTS taxi drivers do not have a specific trip origin and destination. The taxis just exist all over the network, and the current location of an empty taxi may change all the time. In this case, we will also need to determine the pick-up and the drop-off sequence for the rider groups (Agatz et al., 2012).

From October 26 to November 17 in 2006, a field trial of taxi-sharing service was conducted at Taipei Nei-Hu Science and Technology Park. Tao (2007) gave an overview of the taxi-sharing service, presented key algorithms for dynamic rideshare matching processes, described the field trial operation of the system in Taipei Nei-Hu Science and Technology Park and discussed empirical results to provide valuable implications for better taxi-sharing service in the future. A qualitative analysis using the Delphi method (shown in Figure 4) was conducted to survey the degree of satisfaction among Taipei city government, taxi operators, taxi drivers, and passengers. The results revealed that taxi operators were ready to accept Intelligent Transportation Systems (ITS) based technologies because they thought taxi-sharing service would help taxi drivers making more money than before. The passengers were not so satisfied with dynamic taxi-sharing service, for they still hesitate to ride with strangers. Lack of sufficient incentives for taxi-sharing could also discourage passengers from using taxi-sharing services. The sharing fee for each passenger was computed on a distance traveled basis. The final sharing fee for each passenger was computed according to the number of sharing passengers, their O-D data and preferences. The paper did not reveal much detail behind the system, but the results of the numerical tests and the user surveys demonstrated that the outcomes of these heuristic algorithms were fairly plausible. The average matching success rate was 60.3% on the whole. However, the developed algorithms were only applicable to the case of “one-to-many” and “many-to-one.” The case of “many-to-many” which fully

represents dynamic ride matching with any O-D pairs for the taxi-sharing problem was under development at the time (Tao, 2007).

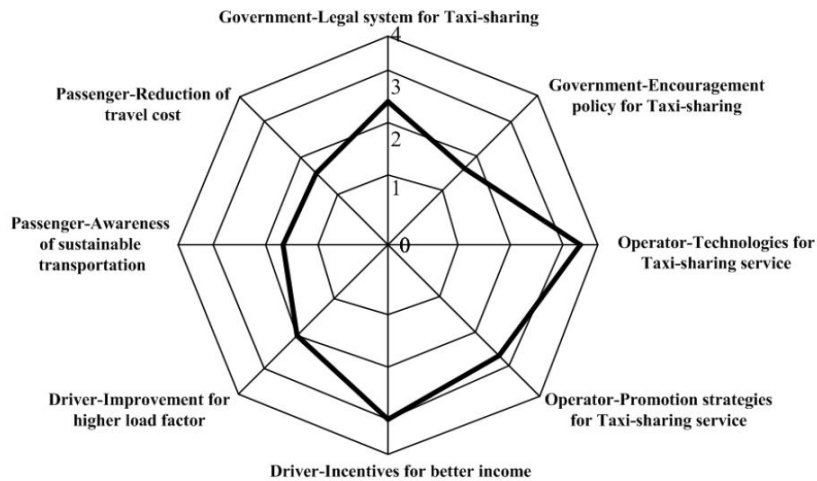
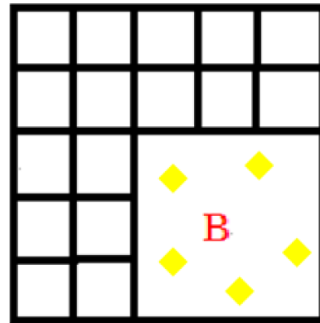


Figure 4 Qualitative Results of the Satisfaction Survey for Taxi-sharing among Government, Operators, Drivers, and Passengers

Chen et al. (2010) proposed a dynamic taxi-sharing system aiming at fuel-saving and pollution-reducing based on ITS technology. Road traffic information was considered when deciding the travel path. Since the main objective of this taxi-sharing system was to save fuel, the primary consideration of the benefit function was the difference in fuel consumption before and after a ridesharing service. As shown in Figure 5, to narrow down the search space for a short response time, they only considered shared-ride taxis whose travel destinations were in the style box of the destination of a ridesharing request (Chen, Liu, & Chen, 2010). This consideration might result in failing to incorporate better taxi candidates into the system outside the style box.

1	2	3	4	5
6	A	8	9	10
11	12	13	14	15
16	17	18	B	20
21	22	23	24	25

We divide our simulation area into 25 subareas.
A,B are source and destination of a ridesharing request, respectively.



For the destination of a ridesharing request(B),
We only consider shared-ride taxis(yellow point)
whose travel destinations are in the style box of B.

Figure 5 Dividing of the Simulation Area

Simulations were used to evaluate the dynamic taxi-sharing system and compare it with other two systems. The results showed that the solution could exactly select a fuel-saving taxi for each ridesharing request and outperform in response time, the number of compared taxis and fuel-saving while compared with existing solutions. (Chen et al., 2010).

Zhang et al. (2013) presented a carpool service, called *coRide*, in a large-scale taxicab network intended to reduce total mileage for less gas consumption. An NP-hard route calculation problem under different practical constraints was formulated. The paper then provided (i) an optimal algorithm using Linear Programming, (ii) an approximation algorithm with a polynomial complexity, and (iii) its corresponding online version. To encourage *coRide*'s adoption, the authors presented a win-win fare model as the incentive mechanism for passengers and drivers to participate. The study

evaluated *coRide* with a real-world dataset of more than 14,000 taxicabs, and the results showed that compared with the ground truth, the service could reduce 33% of total mileage; the win-win fare model could lower passenger fares by 49% and simultaneously increase driver profit by 76%. The paper also pointed out that it is very critical to establish a right policy that would make a large scale deployment feasible (Zhang et al., 2013).

Ma, Zheng, & Wolfson (2013, 2015) proposed a taxi searching algorithm using a spatiotemporal index to quickly retrieve candidate taxis that are likely to satisfy a user query. A scheduling algorithm was then proposed. It checked each candidate taxi and inserted the query's trip into the schedule of the taxi which satisfied the query with minimum additional incurred travel distance. A lazy shortest path calculation strategy was devised to speed up the scheduling algorithm. A mobile-cloud architecture based taxi-sharing system was devised. The service was evaluated using GPS trajectory dataset generated by over 33,000 taxis during a period of 3 months. An experimental platform was built that simulated real user behaviors in taking a taxi. The approach demonstrated its efficiency, effectiveness, and scalability (Ma, Zheng, & Wolfson, 2013, 2015).

Santos & Xavier (2015) dealt with a combinatorial optimization problem that modeled situations of both dynamic ridesharing and taxi-sharing. Passengers and drivers specified their needs, and all drivers defined a price per kilometer. The

problem was to compute routes, matching requests to vehicles in such a way that ridesharing was allowed as long as some restrictions were satisfied, such as the capacity of the vehicle, maximum trip cost of each passenger and maximum delay. Two criteria, maximizing the number of served requests, and minimizing the sum of the costs of all served requests were optimized. The cost-sharing rule divided the cost of each part of the route evenly among the passengers in the car (Santos & Xavier, 2015).

Experiments with instances based on real data were made to evaluate the heuristics and the proposed method. In the simulations with taxis, passengers paid, on average, almost 30% less than they would pay on private rides. (Santos & Xavier, 2015).

Besides academic research, there are existing commercial products with the idea of taxi sharing on the market. UberPOOL from Uber Technologies Inc. matches riders heading in the same direction. Trips are up to 50% less than uberX (a low-cost Uber product). Uber first launched UberPOOL in San Francisco in August 2014, and the product is now available in 29 cities around the world. (Gurley, 2015; Movable Type Scripts; Myhrvold, 2015)

Uber is not alone in launching taxi-sharing services. Its strong competitor Lyft and Sidecar launched similar services - Lyft Line and Sidecar rideshare almost at the same time. Uber does not reveal the percentage of passengers who take UberPOOL over its other products. The company only reveals that "many thousands" take

UberPOOL five days a week to commute to work. While Lyft’s founder Logan Green has said that Lyft Line makes up “the majority” of its rides in San Francisco. In San Francisco, 50 percent of all Lyft rides are now taken with Lyft Line. Sidecar announced that its Shared Rides account for 40 percent of its rides in the cities where it’s launched. (Deamicis, 2015; Wagstaff, 2015)

Although such taxi-sharing products already exist on the market, the matching and routing method behind these commercial products are not revealed, and researchers continuously study and improve the methods.

2.1.4 Summary of the Three Ridesharing Services

Table 2 shows the summary and comparison of the features of the three ridesharing services:

Table 2 Summary of the Features of Three Ridesharing Services

	Traditional Ridesharing	Dynamic Ridesharing	Dynamic Taxi-Sharing
Time aspect	Preset	Dynamic	
Recurring or not	Usually is long-term commuting ridesharing, needs commitment.	Non-recurring; does not need a long-term commitment	

Matching type	Both automated matching and self-choosing matching exist. ^[1]	Automated matching from the centralized system
Cost and billing	Commuting ridesharing participants usually take turns on driving instead of paying each other on each ride. Cost calculating and automated cost calculating also exist.	Automated cost calculating and billing system
Driver O-D pair Features	The driver participants who offer the rides have certain origin and destinations.	Taxi drivers do not have certain origins or destinations.
Route Features	The vehicle always starts from the driver's origin, passing through the riders' origins and destinations, and finish at the driver's destination.	The pick-up and drop-off sequences for all users need to be arranged. The trip ending point is the destination of the last dropped-off user since there is no destination for taxi drivers.
Vehicle ownership requirement	The participants who offer the rides (participate as the driver) must have a vehicle.	Taxi users do not need a vehicle to participate.

Time Restrictions	The drivers and riders usually have time restrictions on their arrival time for the destinations.	The taxi drivers usually do not have time restrictions.
--------------------------	---	---

Notes:

[1].In some traditional ridesharing, there is simply a notice board where participants can post their desired schedule, O-D pairs and contact information. Participants choose ridesharing partners themselves and contact them to make a ridesharing agreement.

2.2 Model Features

We can summarize some considerations from previous ridesharing models below.

2.2.1 Ridesharing Objectives

Most existing studies on ridesharing/taxi-sharing consider one (or a combination) of the following specific objectives when determining ridesharing matches:

- Minimize system-wide vehicle-miles;
- Minimize system-wide travel time;
- Maximize the number of successful matches/matching success-rate;
- Minimize additional incurred travel distance;
- Minimize the maximum cost of each person.
- Minimize the total cost.
- Maximize the total difference in fuel consumption before and after a ridesharing service.

- No systematic objective function.

However, from the taxi providers' perspective, traveling mileage cost is not the only cost, and none of the objective functions above always help the providers gain maximum profit.

2.2.2 Matching Constraints

When proposing matches in a ridesharing system, some constraints on the feasibility of matches must be met. They are listed below:

- The time window of the drivers/rides. For example, Agatz et al. (2011) and Zhang et al. (2013) let a participant specify an earliest possible departure time and latest possible arrival time (N. Agatz, A. L. Erera, M. W. P. Savelsbergh, & X. Wang, 2011; Zhang et al., 2013). Baldacci et al. (2004) and Amey (2011) also allow limits on the actual time that users spend traveling on a given trip. That is, they allow participants to specify the maximum excess travel time (over the direct travel time for their origin to destination) they are willing to accept (Agatz et al., 2012; Amey, 2011; Baldacci, Maniezzo, & Mingozzi, 2004).
- Single-matching constraints need to be considered.
- The vehicle availability and capacity should be considered (Zhang, D. et al., 2013).

- Participants' personal preference could be considered for a higher acceptance rate. For example, female participants may not feel safe sharing a ride alone with a male stranger (Dueker, Bair, & Levin, 1977), while smoking may be another critical issue (Ghoseiri, 2012). Of course, the more restrictions a potential user places on his pool of potential ride-share partners, the more difficult it will be to find successful matches for that user (Dailey et al., 1997).
- Constraints that restrict feasible matches to those that reduce the total travel mileage and individual travel costs need to be considered. Note that not all matches that reduce total mileage can lead to reducing personal travel cost because the cost calculation system may vary.
- Maximum allowable trip cost of each passenger
- Maximum allowable delay for each passenger

2.2.3 Dynamic Strategy

In most of the practical dynamic ridesharing implementations, new riders and drivers continuously enter and leave the system. A driver enters the system by announcing a planned trip and offering a ride, while a rider enters the system by announcing a planned trip and requesting a ride. Drivers and riders leave the system when a ride-share arrangement has been planned and accepted, or when their planned trips “expire.”

Santos et al. (2015) fulfilled the dynamic scheme by using a non-static vehicle set M and requests set N . The set N is empty at the beginning. At each instant of time, new requests may be added to N and matched or expired requests are removed. In the same way, the set M is not static, and it has all vehicles that are available at the current time. To solve this dynamic problem, the day is divided into time periods. For each period, an instance of a static problem is created and solved by a greedy randomized adaptive search procedure (GRASP) (Santos & Xavier, 2015).

Agatz et al. (2011) dealt with this planning uncertainty by using a rolling horizon solution approach. In this approach, the optimization problem to be solved includes all of the offered rides (drivers) and requested rides (riders) that are known at the time of execution and that have not yet been matched.

In a similar setting, Kleiner et al. (2011) applied a rolling horizon solution approach where arrangements are committed as late as possible given the time considerations.

2.2.4 Cost Sharing/Fare Splitting Scheme

The incentive that encourages people to participate in traditional ridesharing or dynamic ridesharing can be reducing personal cost, using the HOV lane, reducing travel time and inconvenience in public transit when they don't have a private vehicle, and social and environmental benefits (Agatz et al., 2012).

In some recurring ridesharing, cost sharing calculation may not be needed when participants take turns driving or when the ridesharing is for completing an experiment. Fagin et al. (1983) proposed a calculation method for “fairness” in the carpool scheduling system, so the system assigns the driver role fairly among all the participants over all times (Fagin & Williams, 1983).

However, when talking about Dynamic Taxi-Sharing, the alternative transport mode is usually non-shared taxi trips. We consider financial incentive as the primary reason why taxi users and drivers would choose DTS instead of non-shared taxi trips. Researchers have been working on developing reasonable methods to share the cost among users.

Geisberger et al. (2010) suggested dividing the cost of the shared part of the trip evenly between the ride-share partners according to *fair sharing rule* in *Algorithmic Game Theory*. (Geisberger et al., 2010; Nisan, 2007)

Agatz et al. (2011) proposed a way to allocate the costs of the joint trip that is proportional to the distances of the separate trips. (N. A. H. Agatz et al., 2011)

Kleiner et al. (2011) proposed an auction-based mechanism to determine the driver’s compensation. Passengers are bidding for increasing their ranking, and thus visibility to drivers, whereas drivers can select passengers according to their preferences. (Kleiner, Nebel, & Ziparo, 2011)

Zhang, D. et al. (2013) raised a term “carpool benefit” B to represent the benefit on the fare of a carpool taxi compared to a non-carpool taxi (Shown in Eq. (1)). They also proposed a method to divide the benefit of the ridesharing to drivers and riders groups according to the percentage of occupied taxi cabs and further divide the benefits among riders’ group according to their non-carpool individual cost. (Zhang et al., 2013)

$$B = \sum_{i=1}^c \tau_i - \tau \quad (1)$$

Where c is the total number of passengers in this carpool; τ_i is the separate non-carpool fare for passenger i ; τ is the regular fare for a distance equal to the carpool distance (not the carpool fare). Thus, the total non-carpool fare of all passengers is given by $\sum \tau_i$, and the regular fare for the carpool distance is given by τ , and their difference is a carpool benefit B (Zhang et al., 2013).

Ma, Zheng, & Wolfson (2013) assumed the following properties for a pricing scheme: (i) taxi fare per mile is higher for multiple passengers than for a single passenger; (ii) the taxi fare of shared distances is evenly split among the riding passengers (Ma et al., 2013).

Denote p as the regular taxi fare per mile, the taxi fare per shared mile is $(\alpha + 1) \times p$. The taxi fare of each passenger can be then automatically calculated by Eq. (2), where d_m is the travel distance shared by m passengers, and c is the capacity of the taxi.

$$fare = p \times (d_1 + \sum_{m=2}^c (\alpha + 1) \times d_m / m) \quad (2)$$

On the other hand, the total fare for all taxi drivers is calculated by Eq. (3), where D_n is the total traveled distance that is not shared and D_r is the total traveled distance that is shared. The appropriate value α is examined to make ridesharing profitable for taxi drivers (Ma et al., 2013).

$$total_profit = p \times (D_n + (1 + \alpha) \times D_r) \quad (3)$$

They also proposed an idea to incorporate a parameter to balance the carpool incentives between the driver and the passengers according to the occupied taxicabs rate (Ma et al., 2013).

2.3 Chapter Conclusion

In this chapter, a comprehensive review of the DTS and relevant studies was presented.

To convince taxi providers to launch a DTS system, we have to prove that the system is beneficial to them. However, none of the existing studies are based on an objective that helps the taxi providers gain the maximum profit. And profit is always the number one consideration for a private company to launch a new product.

On the other hand, the successful launching of a DTS system cannot be completed without enough users. The financial benefit of the taxi fare is crucial for attracting users to a DTS system. Many existing ridesharing services did not focus much on the fare splitting part. To attract more users, we need to design a taxi fare splitting scheme which can fully meet the users' interests.

Besides providing financial benefits for the taxi providers and users, the society should also benefit from a DTS system. As discussed in Chapter 1, the uneven temporal distribution of available taxis is one of the main problems in the current taxi industry. To deal with this issue, we can adopt the idea from Zhang, D. et al. (2013) (Zhang et al., 2013), and design a DTS system with parameters that can self-adjust according to the current taxi occupancy rate, thus encourage or discourage users to

use the DTS service at a specific time period, and to balance the system-wide taxi occupancy rate in real time.

Furthermore, many of the existing ridesharing methods require building up a network before applying the method. The users have to travel from and to the existing nodes in the formulated network. This can limit the application. For a whole city or even larger system, methods that can handle requests with random origins and destinations is more practical.

Thus, a matching method on DTS which can maximize the total provider profit while also attract more users by offering lower fare in the fare calculation scheme can balance the system-wide taxi occupancy rate, and can take random origin/destination requests still needs to be studied.

Chapter 3: Model Formulation

This chapter presents an integer programming formulation for the Dynamic Taxi-Sharing problem. This formulation aims at maximizing the taxi provider's profit while offering taxi service to all the users. Only a portion of the users share taxi rides, and the rest of the users are assigned to regular non-shared taxis.

The basic assumptions of this model are:

- The taxi providers' goal is to maximize the total profit, which is the difference between the total revenue and the total cost while offering taxi service to all users.
- The taxi providers' cost is mainly composed of two parts: fixed cost and routing cost, and each part can be formulated separately.
- The taxi providers' total revenue is proportional to the total taxi fare all the taxi drivers collect (e.g., 80% of the fare is given to drivers, and 20% of the fare is given to the taxi provider company).
- The taxi fare is a linear function of the distance traveled with passengers on board;
- The taxi providers' interest is maximizing the total profit while offering taxi service to all the users who request service;

- All taxi drivers operate according to the taxi provider management center's orders;
- To maintain the advantages of taxi transportation and avoid too much detour and dwell time, each taxi only provides taxi service to at most two taxi user groups (one user group can contain one or two persons but only make one request in the system);
- All taxi requests must be served. The user requests which cannot be matched with other user requests will ride alone;
- Each taxi must finish one assigned trip (dropping off all users) before being reconsidered available in the system for the next assigned trip;
- The number of available taxi user requests is always no greater than twice the number of available taxis.

3.1 Notation

The notations of the model are listed in Table 3, Table 4, and Table 5. Three sets of binary decision variables are used to formulate the model.

Table 3 Data Sets for DTS

Data Set	Description
R^t	The set of taxi users in the system at time window t , $t \in TW$
Q^t	The set of taxi drivers in the system at time window t , $t \in TW$
TW	Time windows set, consists of a set of time window t , $t \in TW$

Table 4 Decision Variables for DTS

Variable	Description
x_{ij}^t	$= \begin{cases} 1, & \text{if rideshare match } (i,j) \text{ is proposed at time window } t \\ 0, & \text{otherwise} \end{cases}$
y_{ki}^t	$= \begin{cases} 1, & \text{if driver } k \text{ is assigned to user } i \text{ at time window } t \\ 0, & \text{otherwise} \end{cases}$
z_{ij}^t	$= \begin{cases} 1, & \text{if the vehicle drops off } i \text{ before } j \text{ at time window } t \\ 0, & \text{otherwise} \end{cases}$

Table 5 Parameters for DTS

Parameter	Description
O_i^t	Origin point of user $i \in R^t$
D_i^t	Destination point of user $i \in R^t$
O_k^t	Origin/current point of taxi driver $k \in Q^t$
$d_{a,b}^t$	The shortest traveling distance from point a to point b at time window t , miles
$t_{a,b}^t$	The shortest traveling time from point a to point b at time window t , hours
Δt	Duration of each time window t , $t \in TW$
M	A large positive value (big-M)
c_f	The fixed operating cost for each taxi, dollars/taxi
c_{pm}	The cost for the taxi provider per mile traveled
$t_{waitingmax}$	The maximum acceptable waiting time for the taxi driver at each taxi users' origin point, hours
$t_{delaymax}$	The maximum acceptable delay time for the taxi riders at each taxi users' destination point, hours

t_{dwell}	Dwell time for picking up a user group, hours;
$d_{detourmax}$	The maximum acceptable detour distance for each taxi user, miles
d_{dmco}	The maximum acceptable detour rate for each taxi user
e_i^t	The earliest allowable departure time for user i at time window t
l_i^t	The latest allowable arrival time for user i at time window t
a_f	The fixed initial fare for every taxi trip, dollars/taxi trip
b_f	The certain fare charge for every mile the taxi user travels, dollars/mile
n_{users}^t	The number of DTS users in the system at time window t
$n_{drivers}^t$	The number of DTS drivers in the system at time window t
ρ	The percentage of the original fare that a DTS user pays; (The discount rate is $1-\rho$)

3.2 Objective Function

Given this setting, we explore the DTS problem in which the taxi providers seek to maximize their profit, which is the difference between the total fare and the total cost.

This objective is also aligned with societal objectives for reducing the total travel distance, fuel consumption, emissions and traffic congestion.

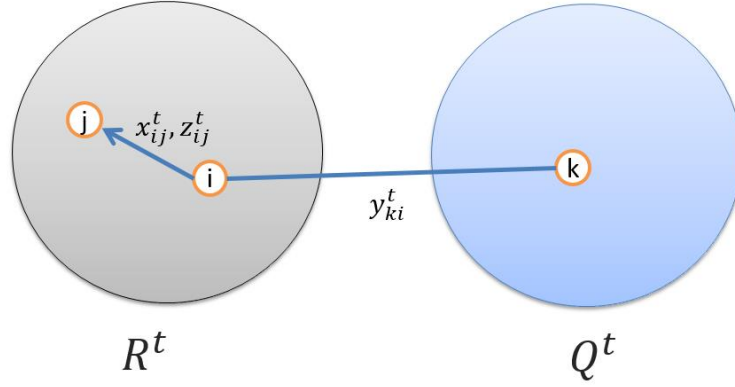


Figure 6 Data Set

We use the rolling horizon approach. The relationship among the data sets and the variables are shown in Figure 6. In each time window t , DTS users i , and j can form a user pair (i, j) . Binary variables are used to represent whether the users will share the same taxi together, given the pick-up sequence (represented by x_{ij}^t), and drop-off sequence (represented by z_{ij}^t), and driver assignment (represented by y_{ki}^t). The taxi users who are not matched will ride alone. The objective function can be written as follows:

$$\begin{aligned}
 & \text{Max Net_Revenue} \\
 & = \text{total_fare} - \text{total_cost} \\
 & = f_{\text{sharing}} + f_{\text{non-sharing}} - C_{\text{fixed}} - C_{\text{routing}}
 \end{aligned} \tag{4}$$

We assume the taxis send GPS location and occupancy status to the central server continuously. For each time window, the optimization is the same, for simplicity, we omit the time window in the formulation and use a rolling horizon approach. The taxi drivers and users who were matched already and accepted the match will leave the system at each time window. Drivers only return to the system as available taxi drivers after the previously assigned users are all completely delivered to their destinations.

The fare in the objective function will be further illustrated in Chapter 4. We consider the total fare consists of two parts: the fare from those sharing taxis and the fare from those non-sharing taxis.

(1) Total sharing fare

We consider this part will be offered a discount rate of $(1 - \rho)$ off their original individual fare (discussed further in Chapter 4). Thus, the total sharing fare can be formulated as:

$$f_{sharing} = \rho \times \sum_{i,j} \left(x_{ij}^t \times \left(2 \times a_f + (d_{o_i, D_i} + d_{o_j, D_j}) \times b_f \right) \right) \quad (5)$$

(2) Total non-sharing fare

The total non-sharing fare can be formulated as (discussed further in Chapter 4):

$$f_{non-sharing} = a_f \times \left(n_{users}^t - 2 \times \sum_{i,j} x_{ij}^t \right) + b_f \times \sum_i \left(d_{O_i, D_i} \times \left(1 - \sum_j (x_{ij}^t + x_{ji}^t) \right) \right) \quad (6)$$

The cost in the objective function for taxi providers mainly contains two parts:

(1) Total fixed cost

We assume there is a fixed cost per trip for each taxi to operate. For each additional taxi trip, the taxi provider has a fixed additional cost.

$$\begin{aligned} total_fixed_cost &= fixed_cost/taxi \text{ (dollars/taxi)} \\ &\times total_number_of_taxis \end{aligned} \quad (7)$$

That is,

$$C_{fixed} = c_f \times \sum_{i,k} y_{ki}^t \quad (8)$$

(2) Total routing cost

We assume the routing cost of a taxi trip is proportional to the traveling distance of the taxi trip. We have:

$$\begin{aligned}
total_routing_cost &= cost_per_mile \text{ (dollars/mile)} & (9) \\
&\times total_traveling_distance \ d_{total} \text{ (miles)}
\end{aligned}$$

The total traveling distance of a DTS trip d_{total} is composed of two parts, the first part is the travelling distance without any passengers, which is the distance from the taxicab's current location to the first taxi rider, and the second part is the travelling distance with one or two passenger groups, which is the distance between picking-up the first rider and dropping-off the last rider. For a DTS trip, when the pick-up sequence is determined, the second part varies according to the drop-off sequence. As shown in Figure 7 and Equation 10, we also take the taxi trips without any sharing matches into consideration in the formulation:

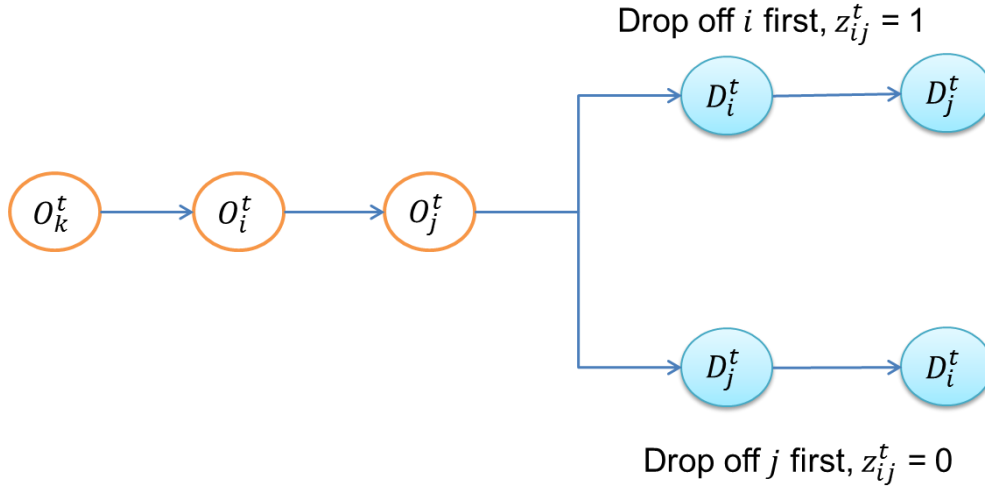


Figure 7 Drop-off Sequence

$d_{total} =$

$$\begin{aligned}
 & \sum_{i,k} (d_{O_k^t, O_i^t} \times y_{ki}^t) \\
 & + \sum_{i,j} \left((d_{O_i^t, O_j^t} + (d_{O_j^t, D_i^t} + d_{D_i^t, D_j^t}) \times z_{ij}^t + (d_{O_j^t, D_j^t} + d_{D_j^t, D_i^t}) \times (1 - z_{ij}^t)) \times x_{ij}^t \right) \\
 & + \sum_i d_{O_i^t, D_i^t} \times \left(1 - \sum_j (x_{ij}^t + x_{ji}^t) \right)
 \end{aligned} \tag{10}$$

The second part of the total distance, the traveling distance of a taxi with passengers

on $d_{totalwp}$ is:

$$d_{totalwp} = \tag{11}$$

$$\sum_{i,j} \left((d_{o_i^t, o_j^t} + (d_{o_j^t, D_i^t} + d_{D_i^t, D_j^t}) \times z_{ij}^t + (d_{o_j^t, D_j^t} + d_{D_j^t, D_i^t}) \times (1 - z_{ij}^t)) \times x_{ij}^t \right) + \sum_i d_{o_i^t, D_i^t} \times \left(1 - \sum_j (x_{ij}^t + x_{ji}^t) \right)$$

Where,

$$z_{ij}^t = \begin{cases} 1, & \text{if } (d_{o_j^t, D_i^t} + d_{D_i^t, D_j^t} - d_{o_j^t, D_j^t} - d_{D_j^t, D_i^t}) \times x_{ij}^t \leq 0 \\ 0, & \text{otherwise} \end{cases}$$

Thus,

$$\begin{aligned} C_{routing} &= c_{pm} \times d_{total} \\ &= c_{pm} \times \left[\sum_{i,k} (d_{o_k^t, o_i^t} \times y_{ki}^t) + \sum_{i,j} \left(\left(d_{o_i^t, o_j^t} + (d_{o_j^t, D_i^t} + d_{D_i^t, D_j^t}) \times z_{ij}^t + (d_{o_j^t, D_j^t} + d_{D_j^t, D_i^t}) \times (1 - z_{ij}^t) \right) \times x_{ij}^t \right) + \sum_i \left(d_{o_i^t, D_i^t} \times \left(1 - \sum_j (x_{ij}^t + x_{ji}^t) \right) \right) \right] \quad (12) \end{aligned}$$

Now we get the total cost for the taxi provider:

$$total_cost = C_{fixed} + C_{routing} \quad (13)$$

Therefore, the objective function above can be further written as:

$$\begin{aligned}
& \text{Max Net_Revenue} \\
& = \text{total_fare} - \text{total_cost} \\
& = f_{\text{sharing}} + f_{\text{non-sharing}} - C_{\text{fixed}} - C_{\text{routing}} \\
& = \rho \times \sum_{i,j} \left(x_{ij}^t \times \left(2 \times a_f + (d_{o_i', D_i'} + d_{o_j', D_j'}) \times b_f \right) \right) + \\
& a_f \times \left(n_{\text{users}}^t - 2 \times \sum_{i,j} x_{ij}^t \right) + b_f \times \sum_i \left(d_{o_i', D_i'} \times \left(1 - \sum_j (x_{ij}^t + x_{ji}^t) \right) \right) - c_f \times \sum_{i,k} y_{ki}^t - \\
& c_{pm} \left[\sum_{i,k} \left(d_{o_k', o_i'} \times y_{ki}^t \right) + \sum_{i,j} \left(\left(d_{o_i', o_j'} + (d_{o_i', D_i'} + d_{D_i', D_j'}) \times z_{ij}^t + \right. \right. \right. \\
& \left. \left. \left(d_{o_j', D_j'} + d_{D_j', D_i'} \right) \times (1 - z_{ij}^t) \right) \times x_{ij}^t \right) + \\
& \left. \sum_i \left(d_{o_i', D_i'} \times \left(1 - \sum_j (x_{ij}^t + x_{ji}^t) \right) \right) \right] \tag{14}
\end{aligned}$$

3.3 Constraints

(1) Single matching for each taxi user constraint

A single matching constraint for users is used to ensure that each taxi user is selected to be included with no more than one proposed match:

$$\sum_j (x_{ij}^t + x_{ji}^t) \leq 1, \forall i \in R^t \tag{15}$$

(2) Driver assignment constraint

Drivers are assigned to the first picked-up user directly. If user j is assigned as the second picked-up user, then there is no driver k assigned directly to user j .

Meanwhile, if a user j is not matched with any users, there is a driver k assigned to the user j directly to provide the individual taxi trip:

$$\sum_k y_{kj}^t + \sum_i x_{ij}^t = 1, \forall j \in R^t \quad (16)$$

(3) The single matching constraint for each driver

Each driver k is selected with no more than one proposed match:

$$\sum_i y_{ki}^t \leq 1, \forall k \in Q^t \quad (17)$$

(4) Single driver assignment constraint for each rider

There is always no more than one driver assigned to rider i , whether or not he or she is matched:

$$\sum_k y_{ki}^t \leq 1, \forall i \in R^t \quad (18)$$

(5) DTS driver guarantee constraint

If a user taxi-sharing pair (i, j) is matched, there must be a driver k assigned to the user pair to provide taxi-sharing service:

$$\sum_k y_{ki}^t \geq \sum_j x_{ij}^t, \forall i \in R^t \quad (19)$$

(6) Taxi users' detour constraint

DTS usually causes a detour for the taxi users. Detour distance is the distance difference between the actual DTS route (the route between being picked up and dropped off for one user) and the non-shared individual taxi trip for the taxi user.

For rider i , we consider both the situation whether he or she is the first or the second picked-up rider:

$$\begin{aligned} detour_i = & \sum_j \left(\left((d_{o_i^t, o_j^t} + d_{o_j^t, D_i^t}) \times z_{ij}^t + (d_{o_i^t, o_j^t} + d_{o_j^t, D_j^t} + d_{D_j^t, D_i^t}) \times \right. \right. \\ & \left. \left. (1 - z_{ij}^t) - d_{o_i^t, D_i^t} \right) \times x_{ij}^t \right) + \\ & \sum_j \left(\left((d_{o_i^t, D_j^t} + d_{D_j^t, D_i^t}) \times z_{ji}^t + d_{o_i^t, D_i^t} \times (1 - z_{ji}^t) - d_{o_i^t, D_i^t} \right) \times x_{ji}^t \right) \end{aligned} \quad (20)$$

We set a maximum acceptable detour distance:

$$0 \leq d_{detouri} \leq d_{detourmax}, \forall i \in R^t \quad (21)$$

Considering the existence of relatively short distance trips, we also set a maximum detour ratio to limit the detour distance within a ratio of the original individual trip:

$$0 \leq d_{detouri} \leq d_{amco} \times d_{o_i, D_i^t}, \forall i \in R^t \quad (22)$$

(7) The constraint for drop-off sequence

For those matched user pairs (i, j) ($x_{ij}^t = 1$), we need to determine the drop-off sequence for the taxi users. The sequence is determined according to the travel distance. The system will choose the drop-off sequence which results in the shorter total travel distance. A decision variable z_{ij}^t is used to represent the drop-off sequence of user pair (i, j) at time window t . If $z_{ij}^t = 1$, the taxi will drop off user i first, otherwise, it will drop off user j first.

$$z_{ij}^t = \begin{cases} 1, & \text{if } \Delta d \leq 0 \\ 0, & \text{otherwise} \end{cases} \quad (23)$$

And $z_{ij}^t = 1$ only exists when $x_{ij}^t = 1$ because we only consider drop-off sequence when the user pair is matched for a DTS trip.

This can be formulated as:

$$\begin{aligned}
M \times (1 - z_{ij}^t) - \Delta d &\geq 0 \\
M \times z_{ij}^t + \Delta d &> 0 \\
(1 - x_{ij}^t) \times z_{ij}^t &= 0
\end{aligned} \tag{24}$$

In which,

$$\Delta d = d_{O_j^t, D_i^t} + d_{D_i^t, D_j^t} - d_{O_j^t, D_j^t} - d_{D_j^t, D_i^t} \tag{25}$$

(8) Taxi waiting time constraint

We set a taxi service delay time constraint for the matched trips. Earliest departure time for taxi user i at time T (T is the time window start time) run is:

$$e_i^t[T] = \max(T, e_i^t) \tag{26}$$

If the taxi arrives earlier than the earliest departure time for the taxi user, there is a taxi waiting time for the taxi provider. The waiting time is the earliest departure time for the user minus the actual taxi arrival time if greater than 0.

$$t_{waitingi} = \max(0, e_i^t - t - \sum_k (t_{O_k^t, O_i^t}^t \times y_{ki}^t)) \times \sum_j x_{ij}^t \tag{27}$$

$$\sum_j \left(\max \left(0, e_i^t - \max \left(e_j^t, t + \sum_k (t_{O_k, O_j}^t \times y_{kj}^t) \right) - t_{dwell} - t_{O_j, O_i}^t \right) \times x_{ji}^t \right)$$

Where t_{O_k, O_i}^t is the travel time between O_k and O_i at time window t .

We set a maximum acceptable waiting time:

$$0 \leq t_{waitingi} \leq t_{waitingmax} \quad (28)$$

We know x_{ij}^t and x_{ji}^t cannot be 0 at the same time. Thus the constraint can be divided into two parts with both part constrained in $[0, t_{waitingmax}]$. The maximum of the two can also be further transformed in solvable format.

(9) Taxi service delay time constraint

We set a taxi service delay time constraint for the matched trips. If the taxi arrives at the destination later than the latest arrival time for the taxi user, the ridesharing cannot be formed. The delay time is the actual taxi arrival time minus the latest acceptable arrival time for the rider if greater than 0. To make it clearer, we write the delay time separately based on whether the user is the first rider or the second rider.

For the first rider,

$$\begin{aligned}
t_{delayi} = \sum_j & \left(\left(\max \left(e_j^t, \max \left(t + \sum_k (t_{O_k, O_j}^t \times y_{kj}^t), e_i^t \right) + t_{dwell} \right. \right. \right. \\
& \left. \left. \left. + t_{O_i^t, O_j^t} \right) + t_{dwell} + t_{O_j^t, D_i^t} \times z_{ij}^t + \left(t_{O_j^t, D_j^t} + t_{dwell} + t_{D_j^t, D_i^t} \right) \right. \right. \\
& \left. \left. \times (1 - z_{ij}^t) \right) \times x_{ij}^t \right) \quad (29)
\end{aligned}$$

For the second rider,

$$\begin{aligned}
t_{delayj} = \sum_i & \left(\left(\max \left(e_j^t, \max \left(t + \sum_k (t_{O_k, O_j}^t \times y_{kj}^t), e_i^t \right) + t_{dwell} \right. \right. \right. \\
& \left. \left. \left. + t_{O_i^t, O_j^t} \right) + t_{dwell} + \left(t_{O_j^t, D_i^t} + t_{dwell} + t_{D_i^t, D_j^t} \right) \times z_{ij}^t + t_{O_j^t, D_j^t} \right. \right. \\
& \left. \left. \times (1 - z_{ij}^t) \right) \times x_{ij}^t \right) \quad (30)
\end{aligned}$$

We set a maximum acceptable delay time:

$$0 \leq t_{delayi} \leq t_{delaymax} \quad (31)$$

In our study, we set $t_{delaymax} = 0$ to eliminate any late arrivals. If needed in future research, this constraint can be relaxed to certain criteria.

The DTS does not add extra delay for the non-matched taxi users compared to a regular taxi trip. We also have the assumption that we will provide taxi service to all the users who request service, whether they are matched or not. We thus omit the waiting and delay time constraint for the non-matched taxi users. Note the taxi users will always have the final option to choose the taxi service or not. After the trip information (driver assignment, matching information, estimated arrival time) is given, the users can choose to accept the service or cancel the request whether they are matched or not.

(10) Nonnegative benefit constraint

For each matched trip, we want to guarantee that the shared fare is no less than the fare calculated by the fare formula based on the distance that the taxi goes on the shared trip:

$$\begin{aligned}
 b_{ij}^t &= \rho \left(2 \times a_f + (d_{O_i^t, D_i^t} + d_{O_j^t, D_j^t}) \times b_f \right) - \\
 &\left(a_f + b_f \times \left(d_{O_i^t, O_j^t} + \left(d_{O_j^t, D_i^t} + d_{D_i^t, D_j^t} \right) \times z_{ij}^t + \left(d_{O_j^t, D_j^t} + d_{D_j^t, D_i^t} \right) \times (1 - z_{ij}^t) \right) \right) \quad (32) \\
 x_{ij}^t \times b_{ij}^t &\geq 0
 \end{aligned}$$

More detailed discussions about the “benefit” will be presented in chapter 4.

3.4 Chapter Conclusion

In this chapter, an integer programming model was proposed for the dynamic taxi sharing problem. A few realistic assumptions were first introduced to help define the problem. The model has an objective function of maximizing the total profit, and a set of realistic constraints to limit the matching sets. The detailed part of the fare formulation is further explained in Chapter 4.

Chapter 4: DTS Fare Calculation Scheme

This chapter aims to design an automated fare calculation scheme for the DTS.

4.1 The Basics

Designing the fare calculation scheme is an essential step in developing the DTS system. We believe that it is reasonable to assume the following features of the DTS fare calculation scheme:

1. the final fare for each DTS user should be less than the fare they pay if riding alone;
2. the summation of the fare from two users, which is also the fare the driver receives, should be more than the fare the driver receives when accomplishing an equivalent distance non-shared trip for a single taxi user.

We assume there is a fixed initial charge a_f for every taxi trip and a certain fare charge b_f for each mile the taxi travels with passengers on board. For simplicity, we omit the slow traffic charge and tolls. The fare can be written as a function of the distance travelled with passengers on board d as follows:

$$f(d) = a_f + b_f \times d \quad (33)$$

By using our DTS system, drivers and riders should both receive a monetary benefit. We assume users get $(1 - \rho)$ discount off their original non-shared fare, that is, they only need to pay ρ of their original non-shared fare $f_{original}$ if they are sharing the taxi trips.

The final fare in DTS for each user should be:

$$f_{user} = \rho \times f_{original} \quad (34)$$

The final fare the driver receives in a DTS trip of two users should be

$$f_{driver} = f_{user\ 1} + f_{user\ 2} \quad (35)$$

The total fare for an equivalent distance non-shared taxi trip should be:

$$f(d_{sharing}) = a_f + b_f \times d_{sharing} \quad (36)$$

Thus, the monetary benefit of one DTS trip for the drivers/taxi providers is:

$$benefit\ B = f_{user\ 1}(d_1) + f_{user\ 2}(d_2) - f(d_{sharing}) \quad (37)$$

$f(d_{sharing})$ is the fare calculated by the current fare system according to the taxi sharing trip distance, which is the fare the driver should receive for offering an equivalent distance non-shared trip for a single taxi user.

$f_{user\ 1}(d_1)$ and $f_{user\ 2}(d_2)$ are the fares calculated by the current fare system according to the individual trip distance, which are the fares the individual users 1 and 2 should separately pay if they ride alone.

4.2 The Dynamics

As we discussed in Chapter 1, the uneven distribution of the taxi availability is the main cause of the current issues. We take the idea from Zhang, D. et al. (2013) as a reference and design our fare scheme to be able to incorporate the dynamic change of the taxi occupancy ratio throughout time to balance this uneven temporal distribution.

We set:

$$\rho = w_f \times \frac{\text{number of empty taxis}}{\text{number of total taxis}} \quad (38)$$

Where w_f is a constant parameter, and $\rho_{min} \leq \rho \leq \rho_{max}$. Figure 8 shows how the parameter could change in different scenarios:

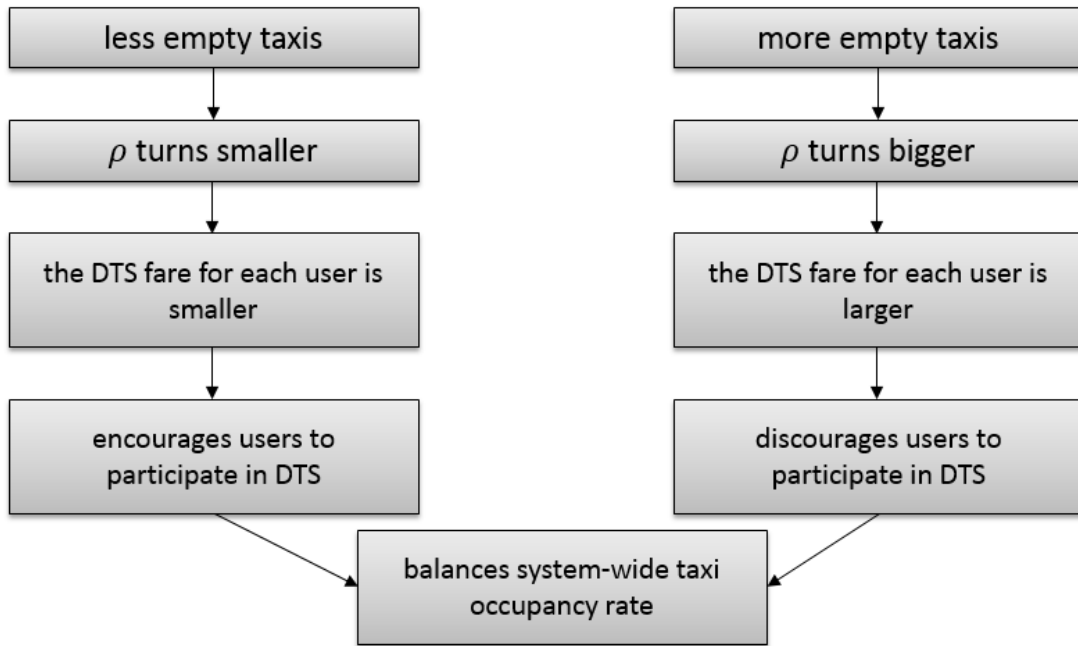


Figure 8 Balance System-wide Taxi Occupancy Rate

4.3 Example

To further illustrate how the system and the fare calculation scheme work, a one-pair DTS matching example is given as follows:

Figure 9, 10, and 11 show one pair of the DTS matching results. The relevant values for x, y, z are $x(30,16) = 1$, $y(44, 30) = 1$, and $z(30,16) = 0$. From $x(30,16) = 1$, we know that user No. 30 and user No. 16 are matched in the DTS system with a sequence of picking up No. 30 first, and No. 16 second. From $z(30,16) = 0$, we know the drop off sequence is dropping off No. 16 first, and No. 30 second. From $y(44, 30) = 1$, we know that driver No. 44 will be assigned to the user pair (30,16).

We can see the two separate trip routes from Figure 9 and Figure 10 according to Google Maps. As shown in Figure 9, the total travel time for an individual trip for user No. 30 is 18 minutes, and the total travel distance is 3.1 miles. As shown in Figure 10, the total travel time for an individual trip for user No. 16 is 11 minutes, and the total travel distance is 1.7 miles.

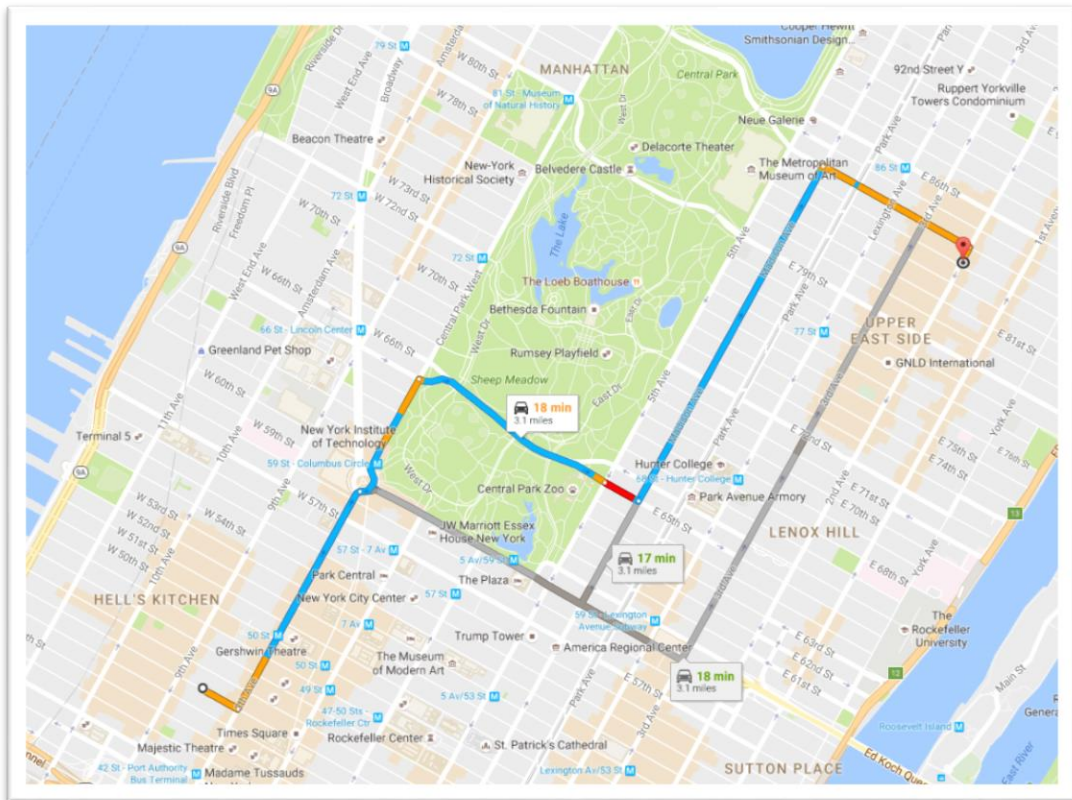


Figure 9 User No. 30 Route from Google Maps

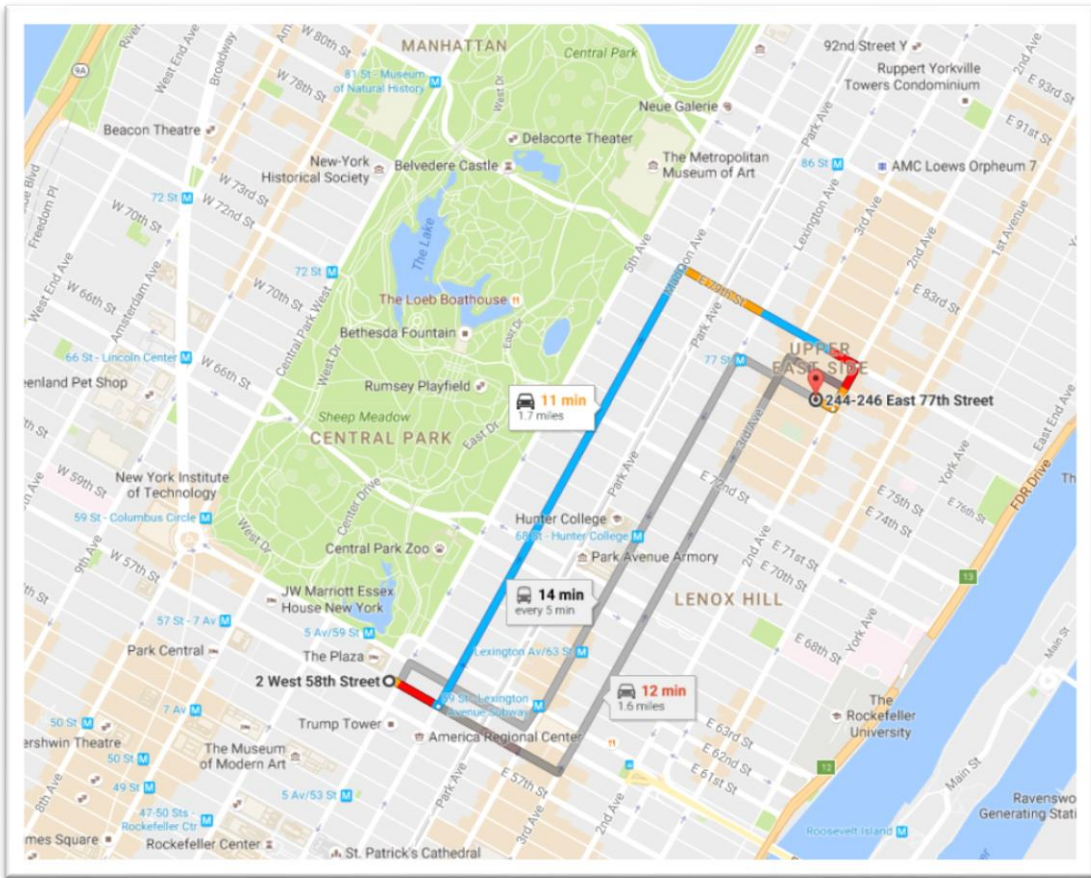


Figure 10 User No. 16 Route from Google Maps

The DTS route is shown in Figure 11, we can see driver 44 passes the origin of user 30, the origin of user 16, the destination of user 16, and the destination of user 30 in this sequence. The total travel time for the shared DTS trip is 25 minutes, and the total travel distance is 3.6 miles.

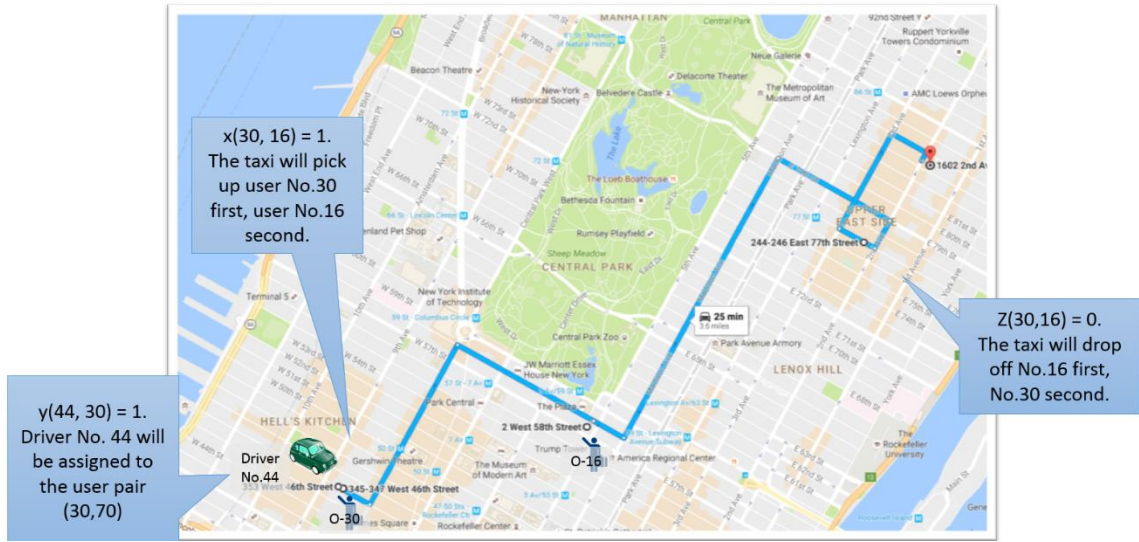


Figure 11 the DTS Trip Route

In the two individual non-sharing trips, if we set $a_f = 2.5$, $b_f = 3$, (the parameter value settings are discussed in Chapter 6, Section 2) the fares can be calculated according to Eq.(39):

$$f_{\text{rider}_1_{\text{original}}} = a_f + b_f \times 3.1 = 11.8 \quad (39)$$

$$f_{\text{rider}_2_{\text{original}}} = a_f + b_f \times 1.7 = 7.6$$

In this shared trip, if we set $\rho = 0.8$, then the fare for each user after the discount is:

$$f_{\text{rider}_1} = \rho \times 11.8 = 9.44 \quad (40)$$

$$f_{rider_2} = \rho \times 7.6 = 6.08$$

Table 6 shows the comparison of the original fare and the discounted fare after the DTS:

Table 6 Fare Comparison before and after DTS for the Two Users

	Distance (miles)	$f(d)$ (\$)	$\rho \times f(d)$ (\$)	Change(\$)
Original Trip 1	3.1	11.8	9.44	-2.36
Original Trip 1	1.7	7.6	6.08	-1.52
Total	4.8	19.4	15.52	-3.88

Now we consider the shared trip. Given the taxi-sharing route is 3.6 miles, the fare for an equivalent 3.6-mile taxi trip can be calculated as:

$$f = a_f + b_f \times 3.6 = 2.5 + 3.6 \times 3 = 13.3 \quad (41)$$

From Table 6, we know the fare of this taxi sharing pair is actually:

$$f_{sharing} = f_{rider_1} + f_{rider_2} = 9.44 + 6.08 = 15.52 \quad (42)$$

Thus, the benefit that the taxi driver gets from offering DTS to this user pair is:

$$f_{sharing} - f = 2.22 \quad (43)$$

The fare comparison before and after DTS for the driver is shown in Table 7:

Table 7 Fare Comparison before and after DTS for the Driver

	Distance (miles)	$f(d)$ (\$)	$f(1)+f(2)$ (\$)	Benefit
DTS Trip	3.6	13.3	15.52	2.22

4.4 Chapter Conclusion

In this chapter, the DTS fare calculation scheme was designed to complete the formulation in Chapter 3, and as an essential structure of the DTS system. An example was given to illustrate how the scheme works. The fare calculation scheme can offer monetary benefits to both taxi users and the drivers (associated with the taxi provider) and can balance the whole network taxi occupancy rate in real time with the self-adjusting parameter.

Chapter 5: A Customized Spectral Clustering

Approach for Preselection on DTS Trips

5.1 Background

In real-world implementations, taxi requests may appear in large amounts, especially during the peak hours. Thus we need an approach which can solve the matching and assignment in a very short time even if the requests set size is large in a time window.

We consider narrowing down the search space for the integer programming model by doing a preselection on taxi trip requests so that the system only needs to consider the trips which have a high possibility to be matched, typically going to the similar direction and in close geographical proximity when calculating a match for the current trip.

To represent the location of a trip, simply using the mean of the latitudes and longitudes of the start and end points does not seem to be representative on a sphere.

We thus consider using midpoint, the half-way point along a great circle path between the two points. The formula to calculate the midpoint is as follows (Movable Type

Scripts):

$$B_x = \cos \varphi_2 \times \cos \Delta\lambda$$

$$B_y = \cos \varphi_2 \times \sin \Delta\lambda$$

$$\varphi_m = \text{atan2}(\sin \varphi_1 + \sin \varphi_2, \sqrt{(\cos \varphi_1 + B_x)^2 + B_y^2}) \quad (44)$$

$$\lambda_m = \lambda_1 + \text{atan2}(B_y, \cos \varphi_1 + B_x)$$

Where φ is latitude, λ is longitude, the subscript 1 and 2 represent the two endpoints, the subscript m represents the midpoint, and Δ represents the difference between the two endpoints. The location of the midpoint on the sphere is shown in Figure 12:

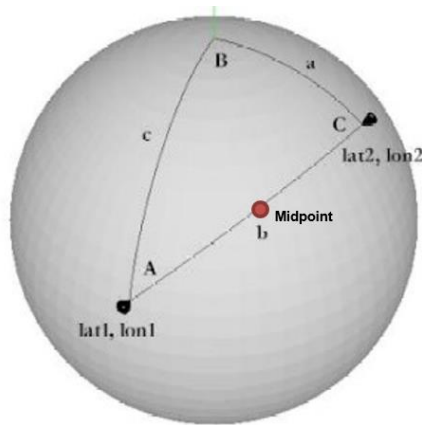


Figure 12 Midpoint

As for the direction of the trip, there are some existing terms representing directions on a sphere. As shown in Figure 13, in navigation, azimuth is a term used for the bearing of a celestial body. Geometrically it is the measure of the arc of the horizon that lies between the elevated pole and the point where the great circle passing through the celestial body cuts the horizon (Kemp, 2005). In this study, we use azimuth to describe the angle to observe the other point from the true north when giving two points on the same surface, that is, the angle between the true north direction and the great circle direction of the two points.

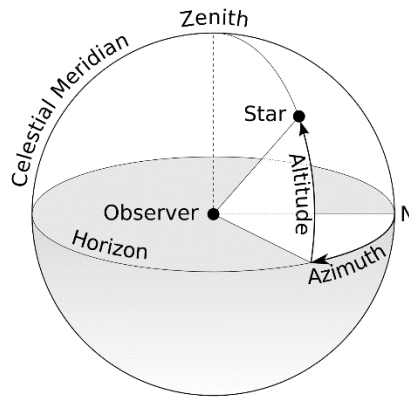


Figure 13 Azimuth in Navigation

The formula to calculate the Azimuth is shown in Eq. (45) (Movable Type Scripts):

$$\theta = \text{atan2} (\sin \Delta\lambda \times \cos \varphi_2, \cos \varphi_1 \times \sin \varphi_2 - \sin \varphi_1 \times \cos \varphi_2 \times \cos \Delta\lambda) \quad (45)$$

Where φ is latitude, λ is longitude, the subscript 1 and 2 represent the two endpoints, and Δ represents the difference between the two endpoints.

In this way, the “preselection” is based on two characteristics: midpoints and azimuth of the trip, which can simply be calculated merely using the trip origin and destination coordinates.

There are some existing concepts that we can use in our “preselection” on the trips in DTS. In data mining, clustering is the process of examining a collection of “points,” and grouping the points into “clusters” according to some distance measure. The goal is that points in the same cluster have a small distance from one another, while points in different clusters are at a large distance from one another. (Leskovec, Rajaraman, & Ullman, 2014)

There are many studies in the transportation field used clustering approach. Shin (2011) proposed a centroid-based heuristic algorithm for the capacitated vehicle routing problem. The method used x and y coordinates to represent point locations in a two-dimensional space and the arithmetic mean of the two coordinates separately to represent the geometrical center of a cluster. The distance between the clusters is represented by the distance between the cluster centers (Shin, 2011).

The study of clustering spatial points has been on for decades, and there are different types of clustering methods. Finding a suitable clustering method is crucial to the problem.

There are two major approaches to clustering – hierarchical and point-assignment. (Leskovec et al., 2014)

1. Hierarchical or agglomerative algorithms start with each point in its own cluster. Clusters are combined based on their “closeness,” using one of many possible definitions of “close.” Combination stops when further combination leads to clusters that are undesirable for one of several reasons. For example, we may stop when we have a predetermined number of clusters, or we may use a measure of compactness for clusters, and refuse to construct a cluster by combining two smaller clusters if the resulting cluster has points that are spread out over too large a region.
2. The other class of algorithms involves point assignment. Points are considered in some order, and each one is assigned to the cluster into which it best fits. This process is normally preceded by a short phase in which initial clusters are estimated. Variations allow occasional combining or splitting of clusters, or may allow points to be unassigned if they are outliers (points too far from any of the current clusters). Common examples of this class are k-means algorithms,

Density-based spatial clustering of applications with noise (DBSCAN), fuzzy clustering, etc.

Spectral clustering uses a similarity matrix as an input and consists of a quantitative assessment of the relative similarity of each pair of points in the dataset. Spectral clustering uses ratio cut, which is related to the sizes of the clusters. Thus, degenerated solutions will be avoided, and the clusters sizes are more likely to be evenly distributed.

5.2 Approach Design

We propose a customized spectral clustering approach for preselection on Dynamic Taxi Sharing based on trip midpoint and azimuth. The basic procedures of the approach are as follows:

1. Update the trip request data set and the driver location data set at the beginning of the time interval;
2. Calculate the midpoints and azimuth for each trip in the given trip request sets;
3. Transfer the azimuth from a pure angle to a normalized vector with (x, y) values (Figure 14);

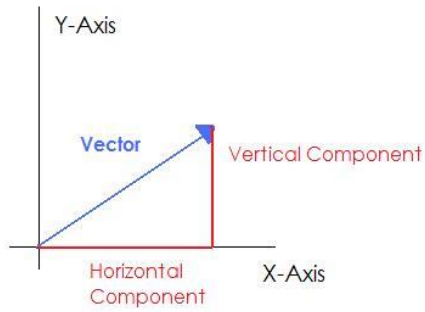


Figure 14 Normalized Vector Pairs

4. Calculate the cosine similarity between the vector pairs;

Cosine similarity, or the cosine kernel, computes similarity as the normalized dot product of X and Y as shown in Eq.(46) (Pedregosa et al., 2011). The Function of $\cos x$ (or cosine similarity value) is shown in Figure 15:

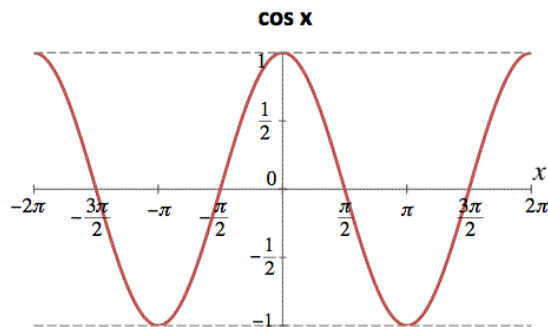


Figure 15 the Function of $\cos x$

$$\text{cosine similarity} = K(X, Y) = \cos(\theta) = \frac{\langle X, Y \rangle}{\|X\| \|Y\|} \quad (46)$$

5. Define and build the “azimuth distance matrix”:

$$\text{Azimuth distance } (i, j) = \begin{cases} M, & \text{if } \cosine\ similarity(i, j) < 0 \\ 0, & \text{if } \cosine\ similarity(i, j) \geq 0 \end{cases} \quad (47)$$

Where M is a large positive number. From Eq. (46) and Figure 15, we know that any two trips with the azimuth angle difference larger than $|\frac{\pi}{2}|$ are assigned a large “Azimuth distance” and thus will be more likely forced into different clusters.

6. Calculate the Haversian distance matrix between trip midpoints for the whole trip data set;

We already talked about using the midpoints to represent the trip locations. In this step, we use Haversian distance formula to build the Haversian distance matrix based on the trip midpoints locations.

7. Design the distance matrix by simply adding Haversian distance we got from step 6 and the Azimuth distance matrix we got from step 5 together:

$$d(i, j) = \text{Azimuth distance } (i, j) + \text{Haversian distance } (i, j) \quad (48)$$

In this way, those trips having azimuth angles larger than 90 degrees are forced to have a very large distance and then be clustered into different clusters.

8. Transform the distance measure to similarity measure:

There are different measurements that we can use to form the similarity measure to perform spectral clustering. We are using a simple exponential function below. Other measurements like Gaussian kernel can also be used.

$$s(i, j) = e^{-d(i,j)} \quad (49)$$

9. Perform spectral clustering using the given matrix.

The approach we followed to perform this step is from the Scikit-learn package (Pedregosa et al., 2011). A second level clustering can also be performed if some specific cluster size is too big for the further calculation. There are several ways to define the number of clusters we use as the input for the clustering. For example, it can be:

$$n_{clusters} = n_{total_requests} // N_{preferred_cluster_size} + 1 \quad (50)$$

Where "//" represents an integer division (floor division), which is the result of the division rounded down to the nearest integer; $N_{preferred_cluster_size}$ is the preferred cluster size; $n_{total_requests}$ is the total number of the taxi requests.

10. Perform the integer programming model in each cluster separately to get the DTS matching and fare calculation results. The calculation in different clusters can be parallel for time consideration. If any of the one clusters cannot be solved in the limited time, we can simply perform a non-sharing taxi driver assignment, or we can do the second level clustering on these clusters and rerun the optimization in those subclusters.

Figure 16 shows the flowchart of the clustering approach:

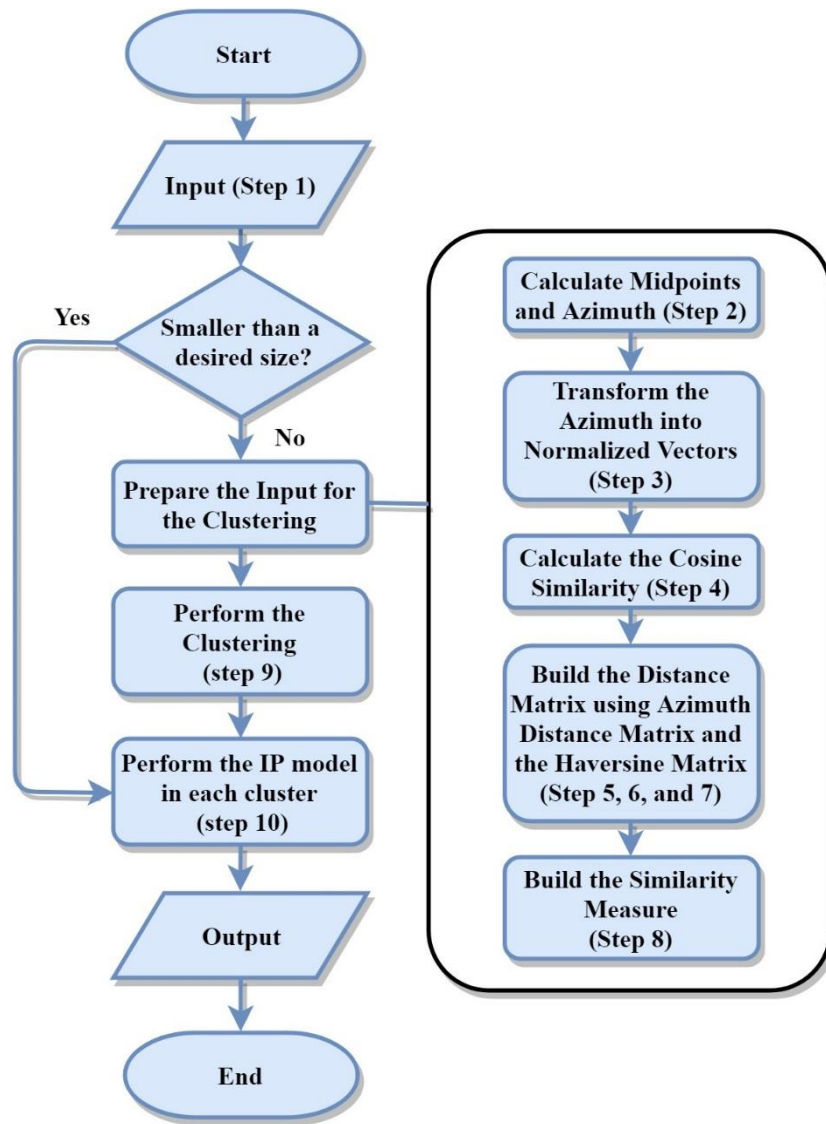


Figure 16 Flowchart of the Approach in One Time Interval

5.3 Chapter Conclusion

This chapter proposed a customized spectral clustering approach for preselection on DTS trips. The approach takes geolocations of the origin and destinations points as input, and considers both taxi trips geolocations and heading directions when performing the clustering. The approach is designed to be conducted at the beginning of each time interval if the trip requests set is larger than a certain size. The performance of the approach will be tested in Chapter 6.

Chapter 6: Case Study

In this Chapter, we implement the model using real-world taxi trip data.

6.1 Data Description and Preparation

We use TLC Trip Record Data (NYC Taxi Limousine Commission, 2016) to implement the model.

This dataset includes trip records from all trips completed in yellow taxis, green taxis, and For-Hire Vehicles (FHV, only available starting 2015) in NYC from 2009 to 2017. Records show that individual taxi trips in the city from January 2009 through June 2015 alone is over 1.1 billion. The data records include fields capturing pick-up and drop-off dates/times, pick-up and drop-off locations, trip distances, itemized fares, rate types, payment types, and driver-reported passenger counts. The data were collected and provided to the NYC Taxi and Limousine Commission (TLC) by technology providers authorized under the Taxicab & Livery Passenger Enhancement Programs (TPEP/LPEP).

Starting July 2016, the latitude and longitude of origin and destination points are substituted by some classified location zone ID. We downloaded December 2015 Boro Taxis (Green Taxis) data which still has the detailed latitude and longitude records. Green taxis cover a much larger pick-up area than those traditional yellow

taxis (as shown in Figure 17) and thus could be more likely to give “general” performance results than the yellow taxis.



Figure 17 Boro Taxis and Yellow Taxi Cover Area

As shown in Figure 17, Boro Taxis can pick up passengers by street hail or a prearranged trip outside the Manhattan exclusionary zone and by prearranged trip only at the airports (NYC Taxi Limousine Commission, 2016).

A 10-row sample of the data set is shown in Tables 8 and 9.

Table 8 Sample of the TLC Trip Record Data Set (Part 1/2)

VendorID	lpep_pickup_datetime	lpep_dropoff_datetime	Store_and_fwd_flag	Rate Code ID	Pickup_longitude	Pickup_latitude	Dropoff_longitude	Dropoff_latitude	Passenger_count	Trip_distance
2	12/1/2015 0:00	12/1/2015 0:00	N	1	-73.9793	40.67369	-73.9909	40.75379	1	7.09
2	12/1/2015 0:00	12/1/2015 0:12	N	1	-73.8442	40.721	-73.8793	40.73832	1	2.26

2	12/1/2015 0:00	12/1/2015 0:13	N	1	- 73.9924	40.694 39	-73.9551	40.7351 2	1	4.91
1	12/1/2015 0:00	12/1/2015 0:02	N	1	- 73.8807	40.748 05	-73.8709	40.7494 7	1	0.6
2	12/1/2015 0:00	12/1/2015 0:05	N	1	- 73.9445	40.714 61	-73.9621	40.7158 2	1	1.17
2	12/1/2015 0:00	12/1/2015 0:16	N	1	- 73.9919	40.690 66	-73.9204	40.6883 4	1	4.04
2	12/1/2015 0:00	12/1/2015 0:23	N	5	- 73.9356	40.833 03	-73.9094	40.8325 9	1	5.53
2	12/1/2015 0:00	12/1/2015 0:17	N	1	-73.955	40.734 04	-73.9876	40.7244 3	1	4.11
2	12/1/2015 0:00	12/1/2015 0:10	N	1	- 73.8073	40.699 49	-73.7786	40.6968 1	1	2
1	12/1/2015 0:00	12/1/2015 0:35	N	1	- 73.9078	40.657 83	-74.1602	40.6207 5	1	24.7

Table 9 Sample of the TLC Trip Record Data Set (Part 2/2)

Fare_amo unt	Extra	MTA_ tax	Tip_amo unt	Tolls_amo unt	Ehail_ fee	improvement_su rcharge	Total_amo unt	Payment _type	Trip_t ype
28.5	0	0.5	5.86	0		0.3	35.16	1	1
11	0.5	0.5	0	0		0.3	12.3	2	1
16.5	0.5	0.5	0	0		0.3	17.8	2	1
4	0.5	0.5	0	0		0.3	5.3	2	1
6	0.5	0	0	0		0	6.5	2	2
15	0.5	0.5	4.08	0		0.3	20.38	1	1
10	0	0	0	0		0	10	2	2
15.5	0.5	0.5	2	0		0.3	18.8	1	1
9.5	0.5	0.5	0.2	0		0.3	11	1	1
66	0.5	0.5	13.46	0		0.3	80.76	1	1

The data dictionary is given in Table 10:

Table 10 Data Dictionary

Field Name	Description
VendorID	A code indicating the LPEP provider that provided the record. 1= Creative Mobile Technologies, LLC; 2= VeriFone Inc.
lpep_pickup_datetime	The date and time when the meter was engaged.
lpep_dropoff_datetime	The date and time when the meter was disengaged.
Passenger_count	The number of passengers in the vehicle. This is a driver-entered value.
Trip_distance	The elapsed trip distance in miles reported by the taximeter.
Pickup_longitude	Longitude where the meter was engaged.
Pickup_latitude	Latitude where the meter was engaged.
RateCodeID	The final rate code in effect at the end of the trip. 1= Standard rate 2=JFK 3=Newark 4=Nassau or Westchester

	5=Negotiated fare 6=Group ride
Store_and_fwd_flag	This flag indicates whether the trip record was held in vehicle memory before sending to the vendor, aka "store and forward," because the vehicle did not have a connection to the server. Y= store and forward trip N= not a store and forward trip
Dropoff_longitude	Longitude where the meter was timed off.
Dropoff_latitude	Latitude where the meter was timed off.
Payment_type	A numeric code signifying how the passenger paid for the trip. 1= Credit card 2= Cash 3= No charge 4= Dispute 5= Unknown 6= Voided trip
Fare_amount	The time-and-distance fare calculated by the meter.
Extra	Miscellaneous extras and surcharges. Currently, this only includes the \$0.50 and \$1 rush hour and overnight charges.
MTA_tax	\$0.50 MTA tax that is automatically triggered based on the metered rate in use.
Improvement_surcharge	\$0.30 improvement surcharge assessed on hailed trips at the flag drop. The improvement surcharge began being levied in 2015.
Tip_amount	Tip amount – This field is automatically populated for credit card tips. Cash tips are not included.
Tolls_amount	The total amount of all tolls paid for the trip.
Total_amount	The total amount charged to passengers (does not include cash tips).
Trip_type	A code indicating whether the trip was a street-hail or a dispatch that is automatically assigned based on the metered rate in use but can be altered by the driver. 1= Street-hail 2= Dispatch

The raw data was filtered before it was used for the analysis. We excluded records of taxi trips with empty GPS locations, and the trips starting and ending at the same location. We also excluded the trips with distances less than 0.4 miles and travel times shorter than 4 minutes since ridesharing will add too much inefficiency for such short trips.

We selected the trip subset of 12/1/2015 which was a Tuesday. The total number of trip requests after cleaning is 41,245. The information we used on each trip is the pick-up time, drop-off time, pick-up and drop-off locations.

We used the same set of drop-off locations to represent taxi drivers' current locations since we do not have a separate data source for taxi locations and the taxis are usually

distributed randomly in the road network. By having the same number of taxi drivers and taxi requests, we can construct the base “non-sharing” model and compare the DTS matching and fare results to the base model. Figures 18 and 19 show the locations of the points on Open Street Map.

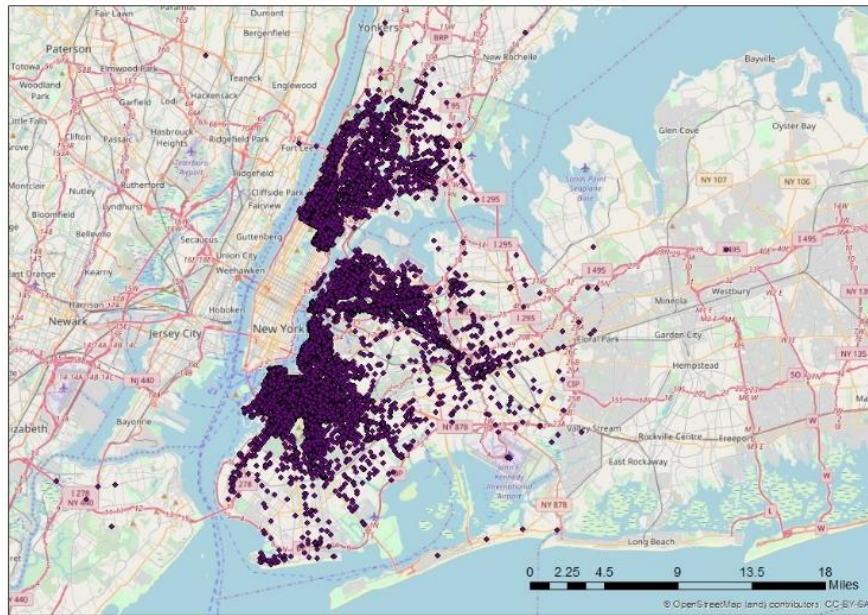


Figure 18 Location of the Pick-up Points

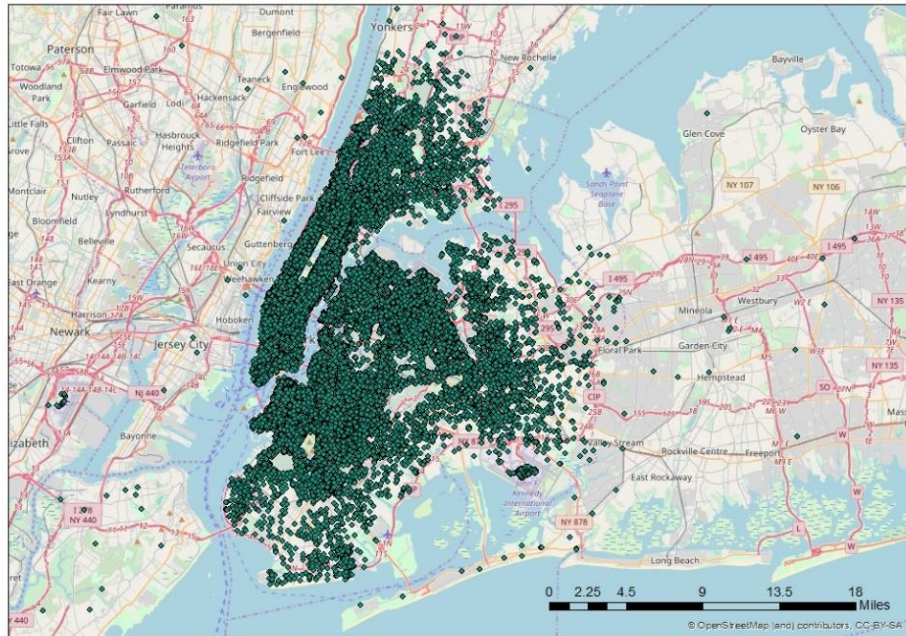


Figure 19 Location of the Drop-off Points

6.2 Implementation Setup

6.2.1 Approach and Environment

The model was implemented in Python 2.7 with Gurobi Optimizer 8.0.0 used as the integer program solver. The computer we used has an Intel Core i5-2400 processor, 3.10 GHz CPU, and 16.00 GB RAM.

We can anticipate that when applied in the real-world, the taxi provider can use more powerful computers and run parallel computation to complete the process thus reduce the total computation time.

6.2.2 Parameters

We use haversine formula (Mwemezi & Huang, 2011; Oxford English Dictionary, 1989) to calculate the great-circle distances between two points on a sphere from their longitudes and latitudes. Figure 20 and 21 show the idea of using haversine formula to calculate the great-circle distance.

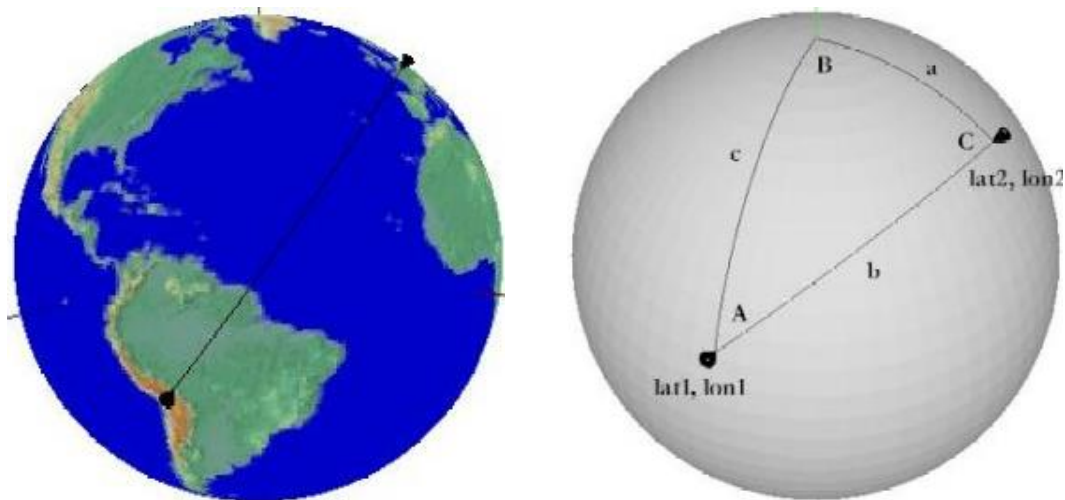


Figure 20 Using Haversine Formula to Calculate the Great-circle Distance

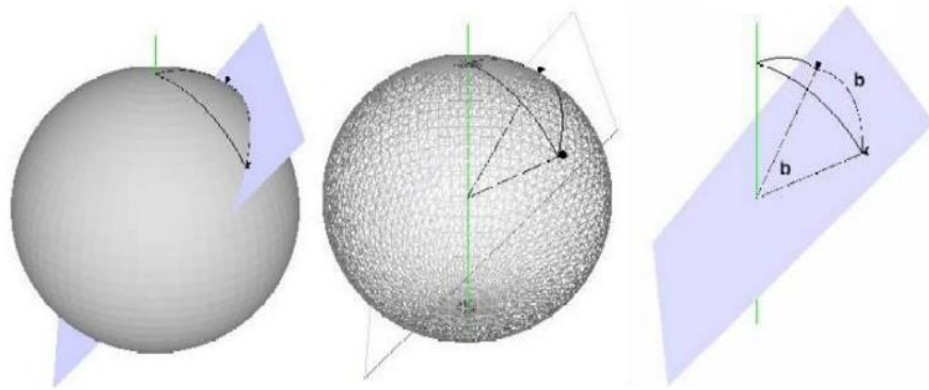


Figure 21: A Rectangular Plane Intersecting a Great Circle Path and the Center of the Earth (Arc b is the Path and the Angular Separation of the End Points).

For any two points on a sphere, the haversine of the central angle between them is given by:

$$hav\left(\frac{b}{r}\right) = hav(\varphi_2 - \varphi_1) + \cos(\varphi_1) \cos(\varphi_2) hav(\lambda_2 - \lambda_1) \quad (51)$$

Where,

- hav is the haversine function:

$$hav(\theta) = \sin^2\left(\frac{1 - \cos(\theta)}{2}\right) \quad (52)$$

- d is the distance between the two points (along a great circle of the sphere),
- r is the radius of the sphere,

- φ_1, φ_2 : latitude of point 1 and latitude of point 2, in radians
- λ_1, λ_2 : longitude of point 1 and longitude of point 2, in radians

On the left side of the equals sign $\frac{b}{r}$ is the central angle, assuming angles are measured in radians.

Solving for b by applying the inverse haversine (if available) or by using the arcsine (inverse sine) function gives:

$$b = r \operatorname{hav}^{-1}(h) = 2r \arcsin(\sqrt{h}) \quad (53)$$

h is $\operatorname{hav}\left(\frac{b}{r}\right)$, or more explicitly:

$$\begin{aligned} b &= 2r \arcsin\left(\sqrt{\operatorname{hav}(\varphi_2 - \varphi_1) + \cos(\varphi_1) \cos(\varphi_2) \operatorname{hav}(\lambda_2 - \lambda_1)}\right) \\ &= 2r \arcsin\left(\sqrt{\sin^2\left(\frac{\varphi_2 - \varphi_1}{2}\right) + \cos(\varphi_1) \cos(\varphi_2) \sin^2\left(\frac{\lambda_2 - \lambda_1}{2}\right)}\right) \end{aligned} \quad (54)$$

We used a mean radius of semi-axes $r = 3958.76$ miles for the radius of the earth in the study (National Imagery and Mapping Agency, 2000).

For real-world taxi trips, the actual travel distance along the route is usually longer than the direct great circle distance due to the real-world road network. We thus

considered using an elevation rate based on the great circle distance to represent the actual travel distance.

To get an approximate elevation rate, we randomly chose 1000 taxi trips from the data set, calculated the great circle distance, and then calculated the ratio by dividing the actual trip distance by the great circle distance for each trip. The average of this ratio is 139.62%. We thus use a 139.62% ratio on the great circle distance to represent the actual travel distance (shown in Table 11).

Table 11 Trip Distance Ratio Calculation Sample

trip_ distance	pickup_ longitude	pickup_ latitude	dropoff_ longitude	dropoff_ latitude	Haversine Distance	Ratio
2.58	-73.9781	40.75249	-73.9786	40.72965	1.578208	1.63476565
4.8	-73.9922	40.72531	-73.923	40.69906	4.056978	1.18314663
0.63	-73.9919	40.7491	-73.9886	40.74295	0.459627	1.37067782
2.51	-73.9903	40.76244	-73.9596	40.77443	1.805745	1.39000808
2.77	-73.9478	40.77634	-73.9767	40.75139	2.291531	1.20879902
1.67	-74.0046	40.73404	-74.0115	40.715	1.363519	1.22477232
1.85	-73.9807	40.74818	-73.9828	40.72815	1.388001	1.33285168
1.41	-73.9789	40.75334	-73.9818	40.76838	1.050013	1.3428401
2.3	-73.9541	40.77477	-73.98	40.75499	1.924815	1.19491987
8.3	-73.9904	40.75654	-73.9392	40.85122	7.06857	1.17421209

There are a number of existing studies on advanced travel time prediction. However, since the travel time prediction is not a focus of this thesis, we used an average speed to calculate the travel time for the sake of simplicity.

We used a similar method to get the average speed (real travel distance/travel time) by dividing the actual travel distance by the actual travel time in the historical data record. The average speed we obtained is 13.83 miles/hour. We thus assumed a constant average vehicle speed of 13.83 miles per hour in the study.

Therefore, we approximated the true travel distances and times and ignored any time-dependency in travel time caused by congestion in the case study.

For the earliest departure time and latest arrival time, we used the actual pick-up time to represent the earliest departure time and assumed a 20 minutes allowable time window for the arrival. That is, we got the latest arrival time by adding 20 minutes to the original trips' arrival time. The parameters values are shown in Table 12:

Table 12 Parameters Values

Parameter	Value
M	1E15
c_f	2
c_{pm}	0.592
$t_{waitingmax}$	0.25
$t_{delaymax}$	0
t_{dwell}	0.017
$d_{detourmax}$	5
d_{dmco}	0.5
e_i^t	The original trip departure time for each trip
l_i^t	The original trip arrival time for each trip +20 mins
a_f	2.5
b_f	3
$N_{preferred_cluster_size}$	30

*Note: Values based on NYC Taxi Limousine Commission charge and AAA: YOUR DRIVING COSTS. How much are you really paying to drive?)

For the fare part, according to NYC Taxi & Limousine Commission, The metered fare information is as follows (NYC Taxi Limousine Commission, 2018):

- ...
- The initial charge is \$2.50.
- Plus 50 cents per 1/5 mile or 50 cents per 60 seconds in slow traffic or when the vehicle is stopped.
- ...

In our study, we omit toll, slow traffic charge, and other charges and thus use a slightly higher parameter for the fare per travel distance than the parameter from the NYC TLC. We keep the initial rate to be the same as the NYC TLC. Thus, Eq. (33) in Chapter 4 becomes:

$$f(d) = a_f + b_f \times d \tag{55}$$

Where $a_f = 2.5$, $b_f = 3$.

6.3 Implementation Results

6.3.1 Preliminary Experiment

We limited the calculation time to 180 seconds and performed the IP model in Chapter 3. The calculation time trend of 119 cases with a size range of [3, 136] is shown in Figure 22. We can see that as the size of the data set increases, the calculation time increases. We considered setting the preferred cluster size at 30 to have better chances of getting the matching results within 180 seconds.

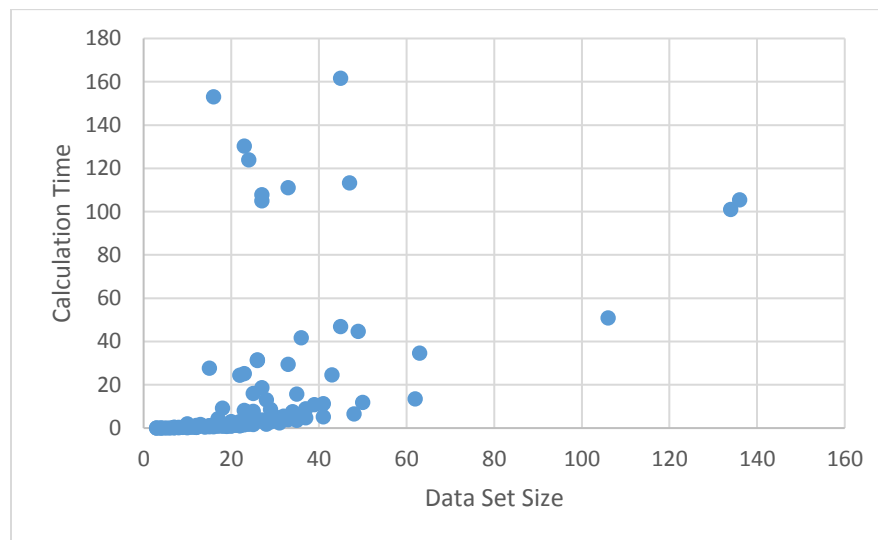


Figure 22 Calculation Time for Different Data Set Sizes

6.3.2 Case Study I

In this case study, we chose three consecutive 30 minutes time intervals during morning peak on a Tuesday:

- 12/01/2015 07:30:00 - 08:00:00
- 12/01/2015 08:00:00 - 08:30:00
- 12/01/2015 08:30:00 - 09:00:00

We updated the users and drivers set at each time period with the corresponding trip set and performed spectral clustering with a reasonable cluster number. Then we ran the optimization separately in each cluster in each time period and compared them with a base model without any taxi-sharing (where we used the same model but forced all the x variables to be zero).

(1) Time Interval 1 (12/01/2015 07:30:00 - 08:00:00)

In this time window, we have a total of 1022 trip requests. The clustering results are shown in Figure 23. The left figure shows the midpoint latitude and longitude, and the right figure shows the azimuth distribution. We can see trips that with close midpoint location and with close azimuth distribution are clustered into the same clusters. (Note the azimuth here is an angle from 0 to 360 degrees, and two trips with an azimuth of 0 and an azimuth of 359 degrees actually have very similar directions.)

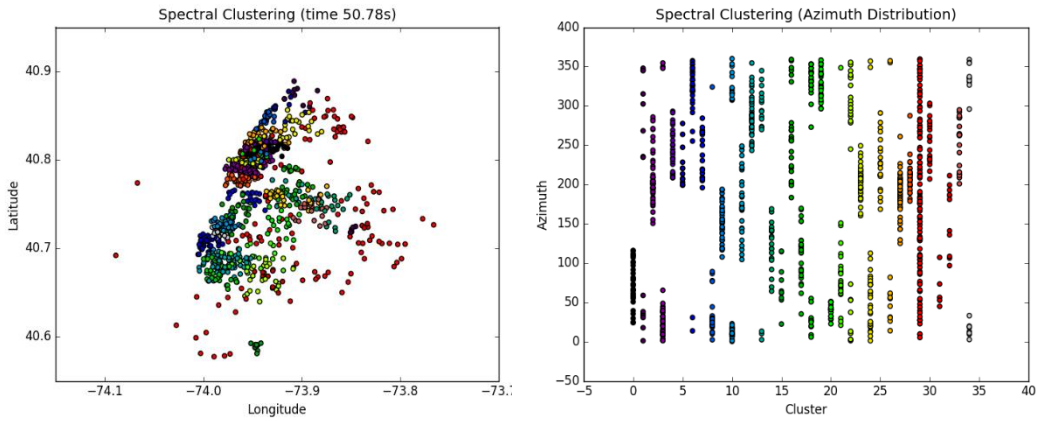


Figure 23 Clustering Results for Time Interval 1

In this time window, we did a sensitivity analysis of the parameter ρ in the optimization step. Results are shown in Table 13:

Table 13 Results for Different ρ Values in Time Interval 1

Name	Non-Sharing Model	$\rho = 1.0$	$\rho = 0.95$	$\rho = 0.9$	$\rho = 0.85$	$\rho = 0.8$	$\rho = 0.75$
Time Limit per Cluster (s)	180						
Number of Trip Requests	1022						
Number of Clusters	35						
Average Cluster Size	29.2						
Number of Clusters w/o DTS Matching Results within Given Time Limit	0	8	7	7	3	2	1
Matched Users	-	638	608	560	668	526	352
Matching Rate	-	62.43%	59.49%	54.79%	65.36%	51.47%	34.44%
Number of Drivers	1022	703	718	742	668	759	846
Number of Drivers Reduction	-	319	304	280	354	263	176
Number of Drivers Reduction Rate	-	31.21%	29.75%	27.40%	34.64%	25.73%	17.22%
Total Profit (\$)	6765.51	8090.83	7790.22	7488.01	7395.69	7139.40	6986.58

Total Profit Increase(\$)	-	1325.32	1024.71	722.50	630.18	373.89	221.07
Total Profit Increase Rate	-	19.59%	15.15%	10.68%	9.31%	5.53%	3.27%
Total Driver Benefit(\$)	-	1816.20	1539.67	1207.49	1284.75	768.77	333.65
Total Distance (Miles)	5979.93	4818.91	4770.02	4474.24	4472.85	4783.42	4978.69
Total Distance Decrease (Miles)	-	1161.02	1209.91	1505.69	1507.08	1196.51	1001.24
Total Distance Decrease Rate	-	19.42%	20.23%	25.18%	25.20%	20.01%	16.74%
Total Distance w. Passengers Onboard (Miles)	3264.88	2925.31	2905.13	2755.18	2804.96	2935.06	3008.00
Estimated Taxi User Acceptance Willingness	-	* (Low Willingness)	**	***	****	*****	***** (High Willingness)

The last row is “Estimated Taxi User Acceptance Willingness.” $\rho = 1$ means that the DTS users will not receive any discount by using this system and thus they will have low willingness in participating the service. As ρ decreases, the users will receive higher and higher discounts and their willingness will be higher.

We can see as ρ decreases from 1.0 to 0.75, the total profit decreases from 8090.83 to 6986.58. This makes sense because we defined $(1 - \rho)$ to be the discount rate that the DTS users receives from the taxi providers if they are matched with another user. The trend is shown in Figure 24:

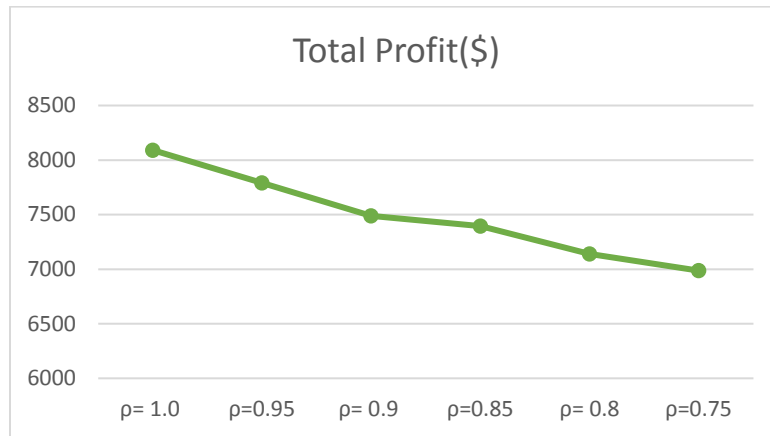


Figure 24 Total Profit for Different ρ Values

We can see the total Driver Benefit has a decreasing trend as the ρ value decreases.

(Shown in Figure 25) From Eq. (37) in Chapter 4 we know this trend is also as expected.

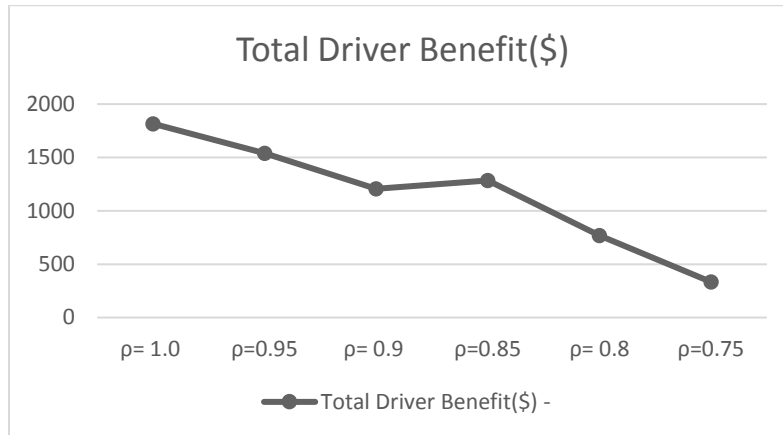


Figure 25 Total Driver Benefit Change for Different ρ Values

Figure 26 shows the trend of “Matching Rate,” “Number of Drivers Reduction Rate,” “Total Profit Increase Rate,” and “Total distance Decrease Rate.” In this case, study,

although $\rho = 1$ gives us the highest “Total Profit”, we have to consider the “User Acceptance Willingness”. Here $\rho = 0.85$ performs well on these rates and maintains a reasonable “User Acceptance Willingness”. We considered using this value for the following case studies.

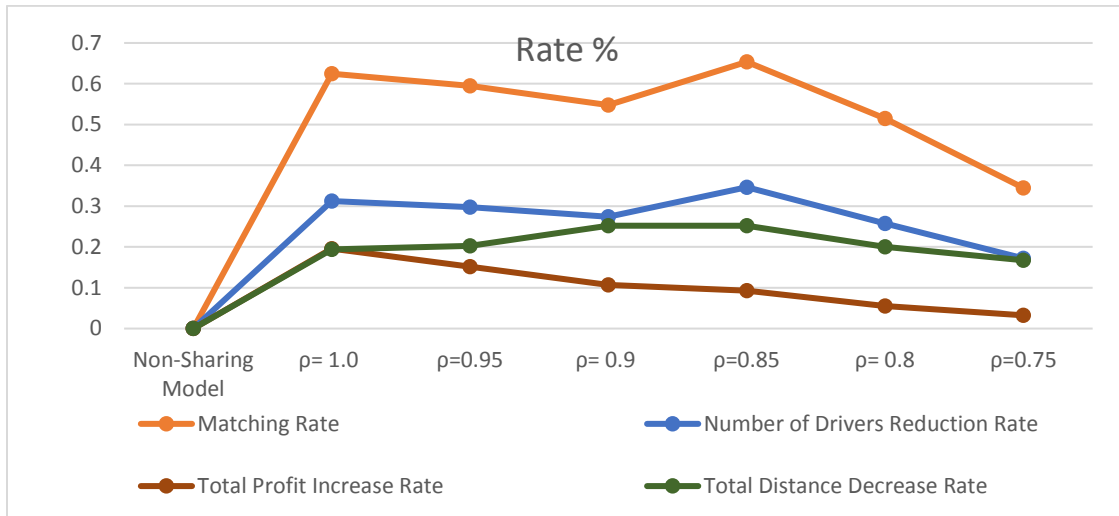


Figure 26 Different Rate Change for Different ρ Values

(2) Time Interval 2 (12/01/2015 08:00:00 - 08:30:00)

There are total 1158 trips in this time interval. We used $\rho = 0.85$, and kept other parameters the same as Time Interval 1. The clustering results are shown in Figure 27:

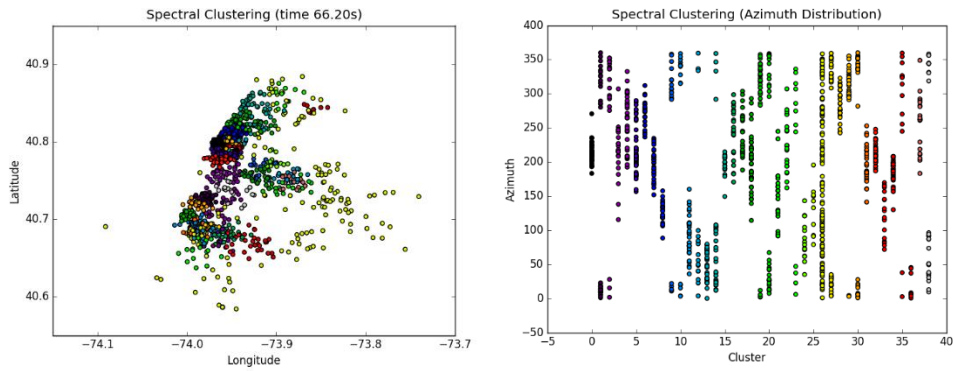


Figure 27 Clustering Results for Time Interval 2

The model results are shown in Table 14:

Table 14 Results for Time Interval 2

Name	Non-Sharing Model	DTS
Time Limit per Cluster (s)	180	
Number of Trip Requests	1158	
Number of Clusters	39	
Average Cluster Size	29.69	
Number of Clusters w/o DTS Matching Results within Time Limit	0	5
Matched Users	-	738
Matching Rate	-	63.73%
Number of Drivers	1158	789
Number of Drivers Reduction	-	369
Number of Drivers Reduction Rate	-	31.87%
Total Profit(\$)	7233.11	7927.10
Total Profit Increase(\$)	-	693.99
Total Profit Increase Rate	-	9.59%
Total Driver Benefit(\$)	-	1300.98
Total Distance (Miles)	6448.87	4853.28
Total Distance Decrease (Miles)	-	1595.59
Total Distance Decrease Rate	-	24.74%
Total Distance w. Passengers Onboard (Miles)	3490.61	3034.92

Estimated Taxi User Acceptance Willingness	-	****
--	---	------

(3) Time Interval 3 (12/01/2015 08:30:00 - 09:00:00)

There are a total 1330 trips in this time interval. We used $\rho = 0.85$, and kept other parameters the same as Time Interval 1. The clustering results are shown in Figure 28:

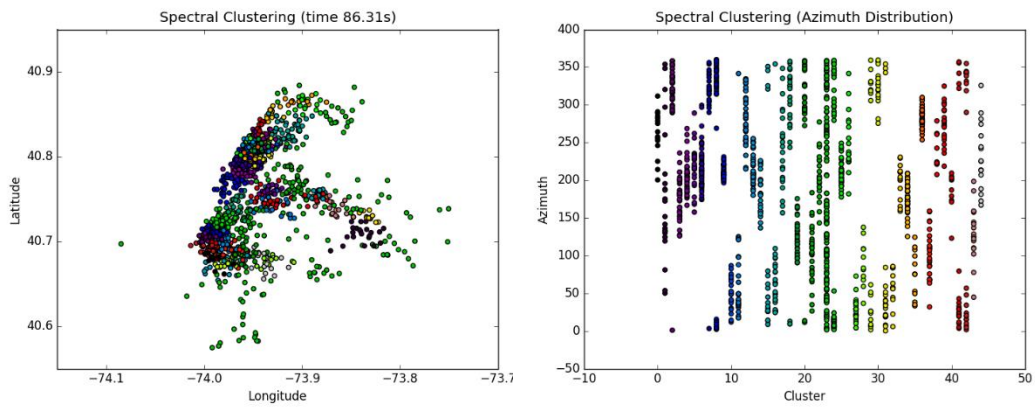


Figure 28 Clustering Results for Time Interval 3

The model results are shown in Table 15:

Table 15 Results for Time Interval 3

Name	Non-Sharing Model	DTS
Time Limit per Cluster (s)	180	
Number of Trip Requests	1330	
Number of Clusters	45	
Average Cluster Size	29.56	
Number of Clusters w/o DTS Matching Results within Time Limit	0	7
Matched Users	-	908

Matching Rate	-	68.27%
Number of Drivers	1330	876
Number of Drivers Reduction	-	454
Number of Drivers Reduction Rate	-	34.14%
Total Profit(\$)	8183.97	10141.86
Total Profit Increase(\$)	-	1957.89
Total Profit Increase Rate	-	23.92%
Total Driver Benefit(\$)	-	2724.60
Total Distance (Miles)	7449.10	5675.64
Total Distance Decrease (Miles)	-	1773.46
Total Distance Decrease Rate	-	23.81%
Total Distance w. Passengers Onboard (Miles)	3976.28	3446.41
Estimated Taxi User Acceptance Willingness	-	****

6.3.3 Case Study II

In this case study, we test the model in three consecutive 10 minute time intervals:

- (1) 12/01/2015 08:00:00 - 08:10:00

There are total 405 trips in this time interval. We use $\rho = 0.85$, and keep other parameters the same as Time Interval 1 in Case Study I. The clustering results are shown in Figure 29:

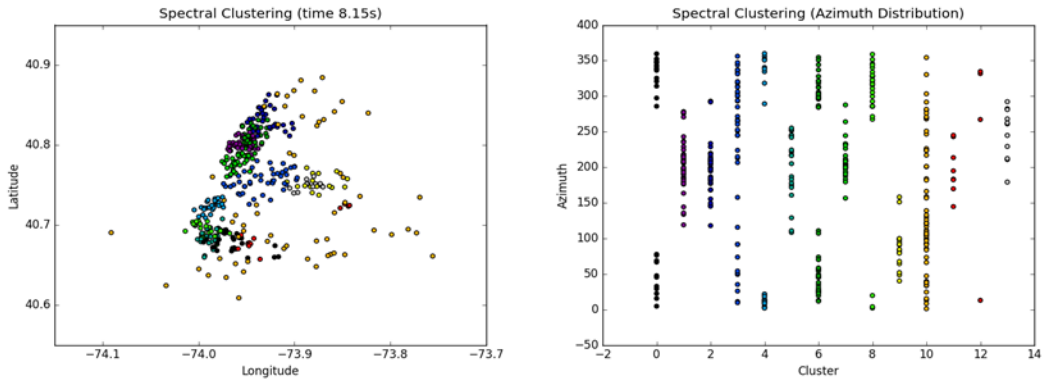


Figure 29 Clustering Results for Time Interval 1

The model results are shown in Table 16:

Table 16 Results for Time Interval 1

Name	Non-Sharing Model	DTS
Time Limit per Cluster (s)	180	
Number of Trip Requests	405	
Number of Clusters	14	
Average Cluster Size	28.93	
Number of Clusters w/o DTS Matching Results within Time Limit	0	4
Matched Users	-	152
Matching Rate	-	37.53%
Number of Drivers	405	329
Number of Drivers Reduction	-	76
Number of Drivers Reduction Rate	-	18.77%
Total Profit(\$)	2622.30	2765.15
Total Profit Increase(\$)	-	142.85
Total Profit Increase Rate	-	5.45%
Total Driver Benefit(\$)	-	193.63
Total Distance (Miles)	2302.99	2022.28
Total Distance Decrease (Miles)	-	280.71
Total Distance Decrease Rate	-	12.19%

Total Distance w. Passengers Onboard (Miles)	1261.06	1201.40
Estimated Taxi User Acceptance Willingness	-	****

(2) 12/01/2015 08:10:00 - 08:20:00

There are total 375 trips in this time interval. We use $\rho = 0.85$, and keep other parameters the same as Time Interval 1 in Case Study I. The clustering results are shown in Figure 30:

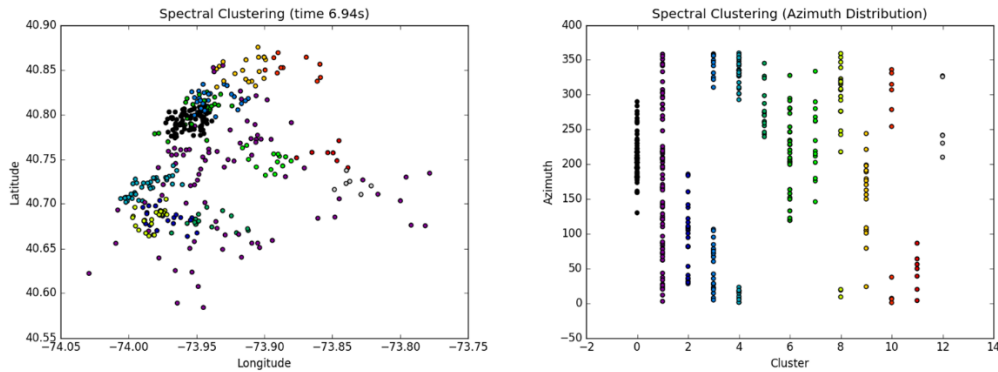


Figure 30 Clustering Results for Time Interval 2

The model results are shown in Table 17:

Table 17 Results for Time Interval 2

Name	Non-Sharing Model	DTS
Time Limit per Cluster (s)	180	
Number of Trip Requests	375	
Number of Clusters	13	
Average Cluster Size	28.85	
Number of Clusters w/o DTS Matching Results within Time Limit	0	5

Matched Users	-	62
Matching Rate	-	16.53%
Number of Drivers	375	344
Number of Drivers Reduction	-	31
Number of Drivers Reduction Rate	-	8.27%
Total Profit(\$)	2308.79	2358.55
Total Profit Increase(\$)	-	49.76
Total Profit Increase Rate	-	2.16%
Total Driver Benefit(\$)	-	79.74
Total Distance (Miles)	1961.98	1852.98
Total Distance Decrease (Miles)	-	109.00
Total Distance Decrease Rate	-	5.56%
Total Distance w. Passengers Onboard (Miles)	1094.56	1068.93
Estimated Taxi User Acceptance Willingness	-	****

(3) 12/01/2015 08:20:00 - 08:30:00

There are total 378 trips in this time interval. We use $\rho = 0.85$, and keep other parameters the same as Time Interval 1 in Case Study I. The clustering results are shown in Figure 31:

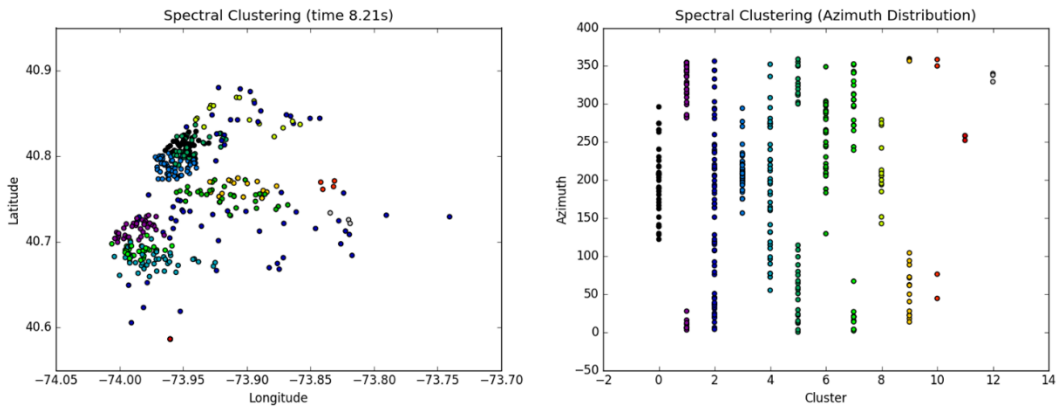


Figure 31 Clustering Results for Time Interval 3

The model results are shown in Table 18:

Table 18 Results for Time Interval 3

Name	Non-Sharing Model	DTS
Time Limit per Cluster (s)	180	
Number of Trip Requests	378	
Number of Clusters	13	
Average Cluster Size	29.08	
Number of Clusters w/o DTS Matching Results within Time Limit	0	4
Matched Users	-	106
Matching Rate	-	28.04%
Number of Drivers	378	325
Number of Drivers Reduction	-	53
Number of Drivers Reduction Rate	-	14.02%
Total Profit(\$)	2405.05	2507.30
Total Profit Increase(\$)	-	102.25
Total Profit Increase Rate	-	4.25%
Total Driver Benefit(\$)	-	145.93
Total Distance (Miles)	2009.86	1797.20
Total Distance Decrease (Miles)	-	212.66
Total Distance Decrease Rate	-	10.58%

Total Distance w. Passengers Onboard (Miles)	1135.30	1087.68
Estimated Taxi User Acceptance Willingness	-	****

We can see the 10-minute time interval case studies generate lower matching rate and less “Total Profit Increase Rate” than the 30-minute time interval case studies. This may result due to the lack of suitable matches in the user pool.

6.3.4 Case Study III

In this case study, we test the “efficiency” of the clustering approach. That is, whether the clustering method can distinguish suitable or unsuitable matches, cluster the suitable matches within the same clusters and maintain relatively good results.

We use a data set from Case Study I Time Interval 1 cluster No.25. This data set has a size of 106, and an original matching rate of 33.96% and the optimal solution can be found in 56 seconds. We conduct another level of clustering and perform the model in the “sub-clusters” to see if the second level clustering can still maintain similar results.

Using the same parameters as before, the second level of clustering clusters the data set into four sub-clusters. Results are shown in Table 19:

Table 19 Results with and without 2nd Level Clustering

Name	Non-Sharing Model	w/o 2nd Level Clustering	w. 2nd Level Cluster	Comparison
-------------	--------------------------	---------------------------------	-----------------------------	-------------------

Time Limit per Cluster (s)	180			-
Number of Trip Requests	106			-
Number of Clusters	1	1	4	3
Average Cluster Size	106	106	26.5	-79.5
Number of Clusters w/o DTS Matching Results within Time Limit	0	0	0	0
Matched Users	-	36	36	0
Matching Rate	-	33.96%	33.96%	0
Number of Drivers	106	88	88	0
Number of Drivers Reduction	-	18	18	0
Number of Drivers Reduction Rate	-	16.98%	16.98%	0
Total Profit(\$)	1227.88	1297.43	1261.53	-35.9
Total Profit Increase(\$)	-	69.55	33.65	-35.9
Total Profit Increase Rate	-	5.66%	2.74%	-2.92%
Total Driver Benefit(\$)	-	68.45	66.51	-1.94
Total Distance (Miles)	821.04	652.90	717.09	64.19
Total Distance Decrease (Miles)	-	168.14	103.95	-64.19
Total Distance Decrease Rate	-	20.48%	12.66%	-7.82%
Total Distance w. Passengers Onboard (Miles)	553.64	523.83	525.18	1.35
Estimated Taxi User Acceptance Willingness	-	****	****	-

We can see we still have the same matching rate in the second level clustering. The “total profit increase” decreased 35.9\$ in the model with the second level clustering, which is a less than 3% decrease rate from the results without second level clustering. Other evaluation terms also changed but not significantly. Overall, we can conclude that the clustering approach can capture the “matching suitability” of the different trips and thus cluster the suitable matches into the same clusters.

6.4 Chapter Conclusion

In this chapter, the proposed model and the clustering approach were tested with real-world taxi data sets. The case studies were based on a few assumptions.

Three sets of case studies were performed. Case study I was based on three 30 minute time intervals; Case study II was based on three 10 minute time intervals; Case study III showed the results with and without a second level clustering to demonstrate the efficiency of the clustering approach.

Results showed that the DTS could increase the total profit, decrease the total number of taxi drivers needed, decrease the total vehicle travel distance, and offer monetary benefits for both users and drivers.

However, the DTS system needs a large enough “user pool” to maintain a preferable matching rate. This is not surprising. According to BuzzFeed News (Anand, 2017), When Uber first launched Pool in San Francisco, just 3,600 of the 35,000 Uber Pool trips completed in the week beginning Sept. 1, 2014, which mean the match rate is just 7.9%. Over time, the number of Pool participants increased, and Uber’s algorithms improved, and the Pool match rate inched higher. Our study still has room to improve over time.

Chapter 7: Summary, Conclusions, and Future Research

7.1 Summary and Conclusions

In this thesis, we proposed a Dynamic Taxi-Sharing system on behalf of the taxi providers. The mathematical formulation of the model was proposed in Chapter 3. The taxi providers can maximize their profit by launching the system. The DTS system can provide taxi sharing solutions to taxi requests by simply taking the geolocations of the current taxi drivers, the geolocations of the users' origins, destinations and desired time window, and does not require to pre-formulate the entire road network as a graph before the application.

We also designed a taxi fare calculation scheme which gives both taxi drivers and riders monetary incentives to use the DTS system in Chapter 4. Besides, the scheme could self-adjust according to the current taxi occupancy rate to balance the occupancy rate over time. This is essential in real-world applications due to the uneven temporal distribution of the available taxicabs.

A customized spectral clustering approach was designed in Chapter 5 to narrow down the search space for the model and served as a preselection procedure before the

implementation of the model. Both taxi route geographical locations and heading directions are considered in the clustering approach.

We performed three sets of case studies in Chapter 6 to validate the feasibility and the effectiveness of the DTS system using real-world taxi trip data (New York City taxi data). Results show that combined with the clustering approach; the model can perform matching and fare calculation in a short time and give monetary beneficial (for both taxi providers and users) matching results.

A sensitivity analysis of the parameter ρ was conducted in the case study I to show how the idea of balancing taxi occupancy in real time works. During peak hours, the taxi provider can reduce the value of the parameter to attract more users to participate in DTS.

Two different time windows were tested. In the 30 minute time interval case, the total profit increase rate ranges from 9.31% to 23.92% for $\rho = 0.85$. In the 10 minute time interval case, the total profit rate ranges from 2.16% to 5.45% for the same ρ value. The difference may be due to the lack of suitable matches in a smaller data set, which means the DTS needs a large enough “user pool” to maintain a preferable matching rate and significant monetary benefits.

Case study III was a test of the efficiency of the clustering approach. Results show that the clustering approach can capture the “matching suitability” among the

different trips and thus cluster the suitable matches into the same clusters for further calculation.

Overall, the proposed model and clustering approach provide considerable monetary benefits to taxi providers, taxi drivers, and taxi users. It was also shown that the maximization of the total profit also aligned with societal objectives for reducing the total travel distance, the total drivers needed, and thus lead to less fuel consumption, emissions and traffic congestion.

7.2 Future Research

There are several venues for future research to improve the model and the solution approach.

First, more efficient heuristic algorithms could be developed and applied to compare with the current method. This is essential to building a real-time implementation. We can also consider separating the user matching and the driver assignment into two steps to reduce the complexity of the problem. That is, in step 1, the user matching could be performed without considering the drivers set, and in step 2, we take the output of step 1 (the user matching results) as the input and perform the driver assignment to the matched user pairs and all the other non-matched users.

Furthermore, the integer programming model itself could be enhanced. We could consider including matching problems with more than two user groups and including

matching for those taxis that already have one user group onboard. More constraints (for example, the riders and drivers' personal preference) can be considered. A thorough sensitivity analysis on the value of different coefficients can be studied in different scenarios.

For further accuracy, we can include the real routing part into the model if we can find an appropriate approach that does not add too much computation time. In this way, we will have the real routing distance and real travel time in the model. The matching could also be adjusted based on the actual route features such as overlapping section ratios.

More methods could be tested in the distance matrix and similarity matrix used for the clustering. And other clustering approaches could also be studied.

Our model still needs a decent user size to have enough suitable matches. Future studies could focus on how to improve the matching rate without bringing down the profit given limited user size.

More studies could be performed on the parameter used to balance real-time taxi occupancy rate and other parameters used in the model. More sensitivity analysis could be added to test the model performance and the parameter settings. Taxi users' behavior, for example, the acceptance rate of an assigned DTS trip according to different discount rates, could also be studied.

For a real-world implementation, the system structure in Chapter 1, Section 3 needs to be further illustrated and tested. The model assumes that the number of available taxi user requests is always no greater than twice the number of available taxis. This may not stand in real-world implementations. The data sets at each time interval should also include a backlog of trips from the previous time interval if they are not yet “expired.”

References

1. Agatz, N., Erera, A., Savelsbergh, M., & Wang, X. (2012). Optimization for dynamic ride-sharing: A review. *European Journal of Operational Research*, 223(2), 295-303. doi:10.1016/j.ejor.2012.05.028
2. Agatz, N., Erera, A. L., Savelsbergh, M. W. P., & Wang, X. (2011). Dynamic Ride-Sharing: a Simulation Study in Metro Atlanta. *Papers Selected for the 19th International Symposium on Transportation and Traffic Theory*, 17, 532-550. doi:10.1016/j.sbspro.2011.04.530
3. Agatz, N. A. H., Erera, A. L., Savelsbergh, M. W. P., & Wang, X. (2011). Dynamic ride-sharing: A simulation study in metro Atlanta. *Transportation Research Part B*, 45(9), 1450-1464.
4. Amey, A. M. (2011). *Real-time ridesharing: exploring the opportunities and challenges of designing a technology-based rideshare trial for the MIT community*. Massachusetts Institute of Technology. Retrieved from Item Resolution URL <http://hdl.handle.net/1721.1/61563> WorldCat.org database.
5. Anand, P. (2017). Leaked Internal Documents Show UberPool Was A Cash-Burning Machine. Retrieved from https://www.buzzfeed.com/priya/uber-pool-burn-rate-frisco?utm_term=.waQVVNbb5Z#.bld664NNgQ
6. Baldacci, R., Maniezzo, V., & Mingozzi, A. (2004). An Exact Method for the Car Pooling Problem Based on Lagrangean Column Generation. *Oper. Res.*, 52(3), 422-439. doi:10.1287/opre.1030.0106
7. Bloomberg, M. R., & Yassky, D. 2014 Taxicab Fact Book. Retrieved from http://www.nyc.gov/html/tlc/downloads/pdf/2014_taxicab_fact_book.pdf
8. Chen, P. Y., Liu, J. W., & Chen, W. T. (2010). A Fuel-Saving and Pollution-Reducing Dynamic Taxi-Sharing Protocol in VANETs. *2010 Ieee 72nd Vehicular Technology Conference Fall*.
9. Dailey, D. J., Loseff, D., & Meyers, D. (1997). *Seattle Smart Traveler* (WA-RD 444.1 TNW 97-08). Retrieved from
10. Deamicis, C. (2015). Uber Says There Have Been Millions of Trips on UberPool, Its Carpool Option. Retrieved from

- <http://www.recode.net/2015/4/16/11561556/uber-says-that-millions-of-people-have-taken-uberpool-its-carpool>
11. Dobrosielski, J., Gray, T., Nhan, A., & Stolen, M. (2007). *Carpool.UMD: community carpooling*. Paper presented at the CHI '07 Extended Abstracts on Human Factors in Computing Systems, San Jose, CA, USA.
 12. Dueker, K. J., Bair, B. O., & Levin, I. P. (1977). Ride Sharing - Psychological-Factors. *Transportation Engineering Journal of Asce*, 103(6), 685-692.
 13. Fagin, R., & Williams, J. H. (1983). A Fair Carpool Scheduling Algorithm. *IBM Journal of Research and Development*, 27(2), 133-139. doi:10.1147/rd.272.0133
 14. Geisberger, R., Luxen, D., Neubauer, S., Sanders, P., Volker, L., the Workshop on Algorithmic Approaches for Transportation Modelling, O., ... Systems, A. L. G. B. R. (2010). Fast detour computation for ride sharing. *OpenAccess Series in Informatics*, 14, 88-99.
 15. Ghoseiri, K. (2012). *Dynamic Rideshare Optimized Matching problem*. Ph.D. Dissertation, the University of Maryland, College Park, MD. Retrieved from <http://hdl.handle.net/1903/13023> WorldCat.org database.
 16. Grynbaum, M. M. (2011). Where Do All the Cabs Go in the Late Afternoon? Retrieved from http://www.nytimes.com/2011/01/12/nyregion/12taxi.html?_r=0
 17. Gurley, B. (2015). Above the Crowd Uber's New BHAG: UberPool. Retrieved from <http://abovethecrowd.com/2015/01/30/ubers-new-bhag-uberpool/>
 18. Hartwig, S., & Buchmann, M. (2007). *Empty seats traveling: next-generation ridesharing and its potential to mitigate traffic - and emission problems in the 21st century*. Retrieved from Technical report, Nokia Research Center:
 19. Huang, C., Zhang, D., Si, Y.-W., & Leung, S. C. H. (2016). Tabu search for the real-world carpooling problem. *Journal of Combinatorial Optimization*, 32(2), 492-512. doi:10.1007/s10878-016-0015-y
 20. Huang, S. C., Jiau, M. K., & Lin, C. H. (2015). A Genetic-Algorithm-Based Approach to Solve Carpool Service Problems in Cloud Computing. *Ieee Transactions on Intelligent Transportation Systems*, 16(1), 352-364. doi:10.1109/Tits.2014.2334597
 21. Kammerdiener, T., & Zhang, H. (2011). Classification of ride-sharing partners based on multiple constraints. *J. Comput. Sci. Coll.*, 26(4), 95-101.

22. Kemp, P. (2005). *The Oxford companion to ships and the sea* (2nd ed. ed.). Oxford [Oxfordshire]; Oxford University Press.
23. Kleiner, A., Nebel, B., & Ziparo, V. A. (2011). *A mechanism for dynamic ride sharing based on parallel auctions*. Paper presented at the Proceedings of the Twenty-Second international joint conference on Artificial Intelligence - Volume Volume One, Barcelona, Catalonia, Spain.
24. Leskovec, J., Rajaraman, A., & Ullman, J. D. (2014). *Mining of massive datasets / Jure Leskovec, Stanford University, Anand Rajaraman, Milliways Labs, Jeffrey David Ullman, Stanford University* (Second edition ed.). Cambridge: Cambridge University Press.
25. Ma, S., Zheng, Y., & Wolfson, O. (2013, 8-12 April 2013). *T-share: A large-scale dynamic taxi ridesharing service*. Paper presented at the Data Engineering (ICDE), 2013 IEEE 29th International Conference on.
26. Ma, S., Zheng, Y., & Wolfson, O. (2015). Real-Time City-Scale Taxi Ridesharing. *IEEE Transactions on Knowledge and Data Engineering*, 27(7), 1782-1795. doi:10.1109/TKDE.2014.2334313
27. Movable Type Scripts. Calculate distance, bearing and more between Latitude/Longitude points. Retrieved from <https://www.movable-type.co.uk/scripts/latlong.html>
28. Mwemezi, J. J., & Huang, Y. (2011). Optimal facility location on spherical surfaces: algorithm and application. *New York Science Journal*, 4(7), 21-28.
29. Myhrvold, C. (2015). It's a Beautiful (Pool) Day in the Neighborhood. Retrieved from <https://newsroom.uber.com/us-california/its-a-beautiful-pool-day-in-the-neighborhood/>
30. National Imagery and Mapping Agency. (2000). Department of Defense World Geodetic System 1984. Retrieved from <http://earth-info.nga.mil/GandG/publications/tr8350.2/wgs84fin.pdf>
31. Nisan, N. (2007). *Algorithmic game theory*. Cambridge: Cambridge University Press.
32. NYC Taxi Limousine Commission. (2016). TLC Trip Record Data. Retrieved from http://www.nyc.gov/html/tlc/html/about/trip_record_data.shtml

33. NYC Taxi Limousine Commission. (2018). Taxicab Rate of Fare. Retrieved from http://www.nyc.gov/html/tlc/html/passenger/taxicab_rate.shtml
34. Oxford English Dictionary. (1989). Cites coinage of term “Haversine” by Prof. Jas. Inman, D. D., in his *Navigation and Nautical Astronomy*. *Oxford University Press*. 2nd ed.
35. Pedregosa, F., Gaël Varoquaux, Alexandre Gramfort, Michel, V., Thirion, B., Grisel, O., . . . Duchesnay, É. (2011). Scikit-learn: Machine Learning in Python. *Journal of Machine Learning Research*, 12 2825-2830.
36. Pieratti, D., Haselkorn, M. P., & Blumenthal, C. (1993). *Bellevue Smart Traveler and cellular telecommunications final report*. Washington, D.C.: Office of Technical Assistance and Safety.
37. Santos, D. O., & Xavier, E. C. (2015). Taxi and Ride Sharing: A Dynamic Dial-a-Ride Problem with Money as an Incentive. *Expert Systems with Applications*, 42(19), 6728-6737. doi:10.1016/j.eswa.2015.04.060
38. Shin, K. (2011). A CENTROID-BASED HEURISTIC ALGORITHM FOR THE CAPACITATED VEHICLE ROUTING PROBLEM. *Computing and Informatics*, 30, 721-732.
39. Smith, A. (2015). U.S. Smartphone Use in 2015. Retrieved from <http://www.pewinternet.org/2015/04/01/us-smartphone-use-in-2015/>
40. Stiglic, M., Agatz, N., Savelsbergh, M., & Gradisar, M. (2015). The benefits of meeting points in ride-sharing systems. *Transportation Research Part B-Methodological*, 82, 36-53. doi:10.1016/j.trb.2015.07.025
41. Tao, C. C. (2007, 21-25 Sept. 2007). *Dynamic Taxi-Sharing Service Using Intelligent Transportation System Technologies*. Paper presented at the 2007 International Conference on Wireless Communications, Networking and Mobile Computing.
42. Turnbull, K. F. (2000). *Assessment of the Seattle Smart Traveler*. U.S. Department of Transportation Federal Transit Administration.
43. U.S. Office of Civilian Defense. (1942). *Car Sharing Club Exchange and Self-Dispatching System*.
44. U.S. Office of Defense Transportation. (1942). *Conservation of Vital War Transportation*.

45. Wagstaff, K. (2015). Riders look for love on UberPool and Lyft Line. Retrieved from <http://sys06-public.nbcnews.com/today/money/riders-look-love-uberpool-lyft-line-2D80511202>
46. Yan, S., Chen, C. Y., & Lin, Y. F. (2011). A model with a heuristic algorithm for solving the long-term many-to-many car pooling problem. *Ieee Transactions on Intelligent Transportation Systems*, 12(4), 1362-1373.
47. Zhang, D., Li, Y., Zhang, F., Lu, M., Liu, Y., & He, T. (2013). *coRide: carpool service with a win-win fare model for large-scale taxicab networks*. Paper presented at the Proceedings of the 11th ACM Conference on Embedded Networked Sensor Systems, Roma, Italy.