Na tomto místě bude oficiální zadání vaší práce

- Toto zadání je podepsané děkanem a vedoucím katedry,
- musíte si ho vyzvednout na studiijním oddělení Katedry počítačů na Karlově náměstí,
- v jedné odevzdané práci bude originál tohoto zadání (originál zůstává po obhajobě na katedře),
- ve druhé bude na stejném místě neověřená kopie tohoto dokumentu (tato se vám vrátí po obhajobě).

Czech Technical University in Prague
Faculty of Electrical Engineering
Department of Computer Science and Engineering

Diploma thesis

Space-Time Demand Adaptive Taxi Routing

Bc. Matej Grajciar

Supervisor: Dr. Malcolm Egan

Study Programme: Open Informatics

Field of Study: Artificial Intelligence

May 11, 2015

Aknowledgements

I would like to thank my supervisor Malcolm Egan for his goodwill and his approach. Moreover I would like to thank my parents for their unconditional support during my whole study.

Declaration

I hereby declare that I have completed this thesis independently and that I have listed all the literature and publications used.
I have no objection to usage of this work in compliance with the act §60 Zákon č. 121/2000Sb. (copyright law), and with the rights connected with the copyright act including the changes in the act.

In Prague 11. 5. 2015 ....

Abstract

Over the last five years, on-demand transport services have begun adopting new approaches, which exploit advanced techniques from artificial intelligence. Most notably Uber have begun to adopt market-based approaches instead of using fixed-pricing. The key feature is that prices are adjusted dynamically with respect to the demand of customers. This means that how much the passengers are prepared to pay matters, in addition to other system parameters also present in traditional on-demand transport systems. Various methods have been proposed to optimize on-demand transport systems, however these are designed for traditional systems, which do not use a market-based approach. One of these methods is a recommendation system to route each taxi to their next location, where they are most likely to obtain a new customer.

In this thesis new, easy to implement, recommendation systems are proposed for market-based on-demand transport systems. The proposed recommendation systems are designed using mathematical models of the system, which leads to novel driver routing algorithms and implementations used for on-line simulation. Moreover these recommendation systems are also evaluated in a realistic system based in the city of the Hague.

Both recommendation systems have different strengths, depending on the size of the operating taxi fleet and the passenger demand. This work provides insights into the right recommendation system that should be used in each scenario.

Abstrakt

V průběhu posledních pěti let začali poskytovatelé přepravy na objednávku využívat nové přístupy, vycházející z pokročilých technik umělé inteligence. Jednou z takových firem je Uber, která zavedla místo klasického fixního nastavení v oblasti taxi služeb dynamické, tržně orientované nastavení ceny. To znamená, že cena přepravy se dynamicky přizpůsobuje poptávce zákazníka. Tento dynamický přístup zavádí nový parametr, který je - na rozdíl od klasického systému - nutno zohlednit, t.j. sumu, kterou je zákazník ochoten zaplatit. Dosud byly navrženy různé metody na optimalizaci přepravy na objednávku, všechny jsou však zaměřeny hlavně na tradičně orientované systémy. Jednou z optimalizačních metod je směrovací systém, který doporučí přepravci lokalitu, kde nepravděpodobněji získá nového zákazníka.

Výsledkem této práce jsou dva nové, snadno implementovatelné směrovací systémy, určené pro tržně orientované přepravy na objednávku. Tyto systémy využívají matematický model popisující reálný přepravní systém, který je následně využit ve směrovacím algoritmu, a tento je pak implementovaný do online simulace. Následně jsou oba tyto směrovací systémy podrobeny experimentu, kde se testuje jejich chování. Každý z těchto dvou systémů má svoje výhody, které se liší v závislosti na počtu operujících vozidel a poptávce zákazníků.

Přínosem práce je poskytnutí přehledu, který ze systémů je vhodnější v závislosti na konkrétní situaci.

Contents

List of Figures

Chapter 1

Introduction

This thesis is focused on market based on-demand transport services such as Uber [34], where journey prices vary over locations in cities and the time of the day. Traditionally, on-demand transport services were dominated by taxis and dial-a-ride [3] services—targeted at the elderly, disabled, and under-served regions. The goal of providers such as Uber [34], Lyft [35] or Liftago [36], is to make cities more accessible by making it more likely for passengers to obtain a ride and also allow drivers to generate more business.

Due to the rising popularity of Uber, there is widespread interest in on-demand transport marketplaces, where dynamic pricing is exploited to balance passenger demand and driver supply. As such, efficient on-demand transport marketplaces must account for how demand and supply varies both temporally and spatially within a city, and involves design of pricing mechanisms, matching passengers to drivers, and planning routes of the drivers. Crucially, the long-term profit of the service provider, the short-term profit of the drivers, and the proportion of passengers that are served are all key factors that influence the design.

The focus of this thesis is to model, design algorithms for, implement, and evaluate routing recommendation systems into an agent-based model of a on-demand taxi service market mechanism. The purpose of this routing recommendation system is to direct a taxi driver towards the next profitable customer, which decreases the waiting time for the taxi and increases the driver's profit.

The main motivation is to improve these market based on-demand transport services from the perspective of passengers,drivers and the provider. A similar approach has already been proposed and successfully proven viable for a fixed-priced taxi services. In this thesis a novel *profit-aware potential* approach of the routing algorithm for an on-demand, is proposed. It considers not only the demand of the service but also the expected profit of the customer. This point of view, ensures not only meeting the customer's requirements on the transport service, but also satisfies the driver's personal objectives of maximizing the profit.

The recommendation systems account for the prices the passengers are prepared to pay and also the cost of the fares, which play an important role in the market-based on-demand transport systems.

In this thesis two recommendation systems are proposed. One *customer-aware*, that takes into account the placement of taxis over location, the expected arrivals of customers and the cost of routing. This system is aimed for the traditional on-demand transport services.

Second recommendation system, *profit-aware*, is designed for the marked-based on-demand transport services. It means that it takes into account in addition to the parameters used in the *customer-aware* recommendation system, also the price that are passengers prepared to pay, the possibility that customers will refuse and the cost of the journey.

This thesis shows that using a recommendation system and solving the routing problem, represents a great asset for the on-demand transport service. The experiments have shown that the performance of the two proposed recommendation systems depend on the size of the operating taxi fleet, the passenger demand and given system model.

This thesis is composed following:

1. **Background and related work** Chapter 2: where the basic concepts of the thesis with related literature are reviewed.

2. **System Model** Chapter 3: where the agent-based model is is proposed and it is introduced the market mechanism inspired bu Uber's approach.

3. **Proposed Approach** Chapter 4: where the mathematical formulation of the profit-aware recommendation algorithm is introduced. A recommendation algorithm for customer-aware (traditional), on-demand transport services, that is based on the related work, is proposed 4.0.4. A Novel recommendation algorithm for profit-aware (novel and so-called smart) taxi providers is proposed, see Section 4.0.3.

4. **Implementation** Chapter 5: where the both the algorithms were implemented into novel agent-based simulation model (Mobility Testbed), to create a recommendation system. Furthermore the algorithm's implementation details are discussed.

5. **Experiments** Chapter 5: where the performance of both the approaches are tested and evaluated and the results are analysed in detail.

6. **Conclusion** Chapter 7: where the experimental results are summarized and discussed in a broader picture.

This material was submitted to the 18th IEEE International Conference on Intelligent Transportation Systems, held on September 15-18, 2015 [37].

Chapter 2

Background and related work

The purpose of this chapter is to provide a general overview of the background for agent-based modeling and market mechanisms. This material forms a basis for the taxi placement problem that is solved by the recommendation system proposed in this thesis.

Following the background overview, the state of the art algorithms, that have been proposed to solve the driver placement problem, are surveyed.

2.0.1 Agent-based Modelling

On-demand transport systems are inherently agent based modelled, because it has the ability to simulate each self interested participator individually. Moreover, the decision-making process is itself different from passenger to passenger as well as from driver to driver. An agent-based model consists of different groups of agents and each of this group is performing their own type of actions according to their goals, abilities and the rules of the system. The aim of this technique is to simulate as precisely as possible the real word behavior, in this case the taxi service transport [1].

2.0.2 Market mechanism

This term represents the exchange of goods or services (taxi service) for money between the buyers (customers) and the seller (provider of the service). The aim of the market mechanism in relation with on-demand transport is to create a well designed system of trade to provide as best service distribution as possible. Which would lead to both customer satisfaction of the demand and providers satisfaction of the gained profit.

There are many types of market mechanisms that deals with allocating the resources. Such as various types of auctions, where buyers compete against each other in order to win the goods, and posted price mechanisms,where the seller generates an offer which can be refused or accepted by buyers [5]. These have been widely applied in applications of artificial intelligence; e.g., for finding an equilibrium in allocating resources. On-demand transport markets are two-sided as both passengers and drivers must be negotiated with by the provider. There are several possible approaches; e.g., double auction [7], Second-price sealed-bid auction etc.

2.0.3 On-demand transport market mechanism

Market mechanism relies on the agent-based modelling since it itself implies that both sellers and buyers are self interested with their own intentions and goals. So that, in this thesis is the market mechanism being applied on the agent-based on-demand transport model. Note that not all on-demand transports use a market mechanism. The on-demand transport is a form of public transport, not operating in regular intervals but irregularly only when requested. It is able to cover specific requests of customers and it is able to cover the demand when regular transport is insufficient [29]. On-demand transport, especially taxis, are nowadays a very discussed market, especially with omnipresent smart technologies. Customers as well as drivers themselves can be tracked down via GPS and this information represents a major contribution for the operating companies. They are able to predict and optimize the taxi service for both themselves, in terms of profit and costs, and also for the passengers. New companies in the transport market, such as Uber, Lyft or Liftago [34, 35, 36], incorporate the smart technologies into the taxi service to create a mobile-app-based transportation network. This novel approach has a huge potential in terms of improving the service for customers and drivers as well as in reducing the overall resources and emission, for example by ride-sharing [2]. Using such service the provider is able to get an estimate price for the taxi fare before you finish, the customer can track down the taxi drivers that is supposed to pick him up, and overall it presents a transparent way of taxi business. Among the price transparency it also offers safety, all the drivers are registered and fares are stored, in case of accident or need of backtracking someones position.

A non-fixed pricing means that the prices for same provided service can differ in relation to the current demand of this service depending on time, weather and more [4]. As known the conventional taxi companies often have fixed prices for flag drop, waiting and per kilometer. The on-demand market mechanism provides an opportunity for the small companies that are not able to compete with other companies through the cost, which is the only option to increase their profit with a fixed-priced rate per kilometer in the taxi service [30]. This allows the new based companies, with novel approaches, to start business among long established companies.

Depending on the supply and demand in the transport market some companies adjust their prices accordingly and use an user-oriented non-fixed pricing called the *surge pricing* [39]. In the case of thigh demand the number of available cars gets tight and during this time, dynamic pricing will automatically go into effect to encourage more drivers to get out on the road and help ensure users always have a ride when they need it most. As demand goes down and more cars free up, the rates will go back to normal.[34]

Surge pricing is still a discussed topic, especially during peaks and events such as New Year's Eve the prices can rise up to seven times [28]. On the other hand this kind of pricing allows the customer to decide if he accepts or rejects the price, because it is known in advance. So even if the price is too high for a customer and he is in deep need of a taxi he can get one in spite of the regular services when getting a taxi during the demand peaks is uncertain.

The crucial task of an on-demand market mechanism is to find the equilibrium among customers and operating taxis. The model should be able to cover the customers request depending on the demand [31].

A key feature of these market based approaches is that prices (both for passengers and drivers) are dynamic, depending on the passenger demand and the supply of drivers. These prices in turn affect key system performance metrics, including the proportion of passengers served, the daily profit of the drivers, and the average long term profit of the service provider. Due to the crucial role that driver supply plays, ensuring that there are neither too many nor two few drivers in each location can significantly improve the performance of the system. As such, selecting the location that each driver waits for a passenger is an important system design problem. This driver placement problem is also non-trivial as although the placements are suggested by the provider, each driver is self interested, which means that the placement must be desirable for him and not only for the system as a whole.

2.1 Related Work

Finding supply and demand equilibrium is crucial for every market. To find this equilibrium and thus successfully cover the demand of the customers, the placement of the taxis in the space and time is the most important task for every provider. Importantly the improvement can be achieved in various fields of the market, the state of art approaches focuses on the after drop-off routing problem.

Much research has focused on modelling the demand curve, to provide as best as possible prediction of the demand, others tend to prefer to use this solid but not perfect estimation to generate a routing path for the drivers, that could in the end lead to better overall performance than simple scheduling with imperfect information about the system.

The work [15] focuses on predicting human mobility from discovering patterns in the number of passenger pick-ups quantity from urban hotspots, while using a large-scale real-world data from the taxi's GPS sensors. Importantly it considers the effect of the weather (considers forecast) on the spatial-temporal variation of passengers in a hotspot. The prediction shows an error of only 5.8%, and their guidance system provides an improvement in the time taken and distance travelled, to find their next passenger, by 37.1% and 6.4%, respectively. Although this paper provides a very good prediction model with a solid improvement for the drivers while using their algorithm, this approach lacks the overall view of the market mechanism, because it only considers the passenger pick ups. Neither the personal interest of the passengers nor driver is considered.

The analysis of efficient and inefficient passenger-finding strategies were proposed in [10], based on the pattern analysis of top- and ordinary-performance taxi group GPS dataset. This approach uses a large-scale model, while examining the features in the top-performance taxi group, in order to translate their strategy into the machine-understanding formalism, to be able to provide it to ordinary-performance drivers and thus improve the overall efficiency of the service. This approach considers the time and the location and it is solely focused on the efficiency, which means the profit of the customer is not accounted for.

The "T-finder" recommender system for driver placement, proposed in [14], was designed to lower the energy consumption of the taxis by reducing the unnecessary cruising of the *vacant* taxis, in search for their next customer. This approach considers along with the pick-up position also the drop-off position and the length of the trip, which is a significant aspect of generated profit, that is the most important interest of the driver. Moreover, the T-finder" method, can precisely detect the parking places, from the GPRs data, with a precision over 90%, thus it is able to provide a guiding system also for the passengers towards these parking spots, while considering the waiting time in each of them, in order to get the taxi as soon as possible. This work is focused on covering the demand with higher efficiency, from the perspective of the passenger in lowering the waiting time for the taxi and also from the perspective of the driver by lowering their driving costs. Despite these positives and proved benefits of this approach, it does not consider various pricing of the passengers nor the possibility they can refuse, i.e. is focused on the fixed-price taxi service model in Beijing.

The need on finding the supply and demand equilibrium is closely described in the paper [9], which proposes a framework, that can provide effective insight into the spatio-temporal distribution of taxi–passenger demand for a 30-min horizon. This paper focuses on the real-time choice problem of which is the best taxi stand (location) to go to after a passenger drop-off . It considers the passenger demand over time, the expected revenue, cost and number of taxis in each stand. In the paper is presented a model for predicting the number of services that will emerge at a given taxi stand. The model has been correctly predicting the 506 873 tested services with an aggregated error measurement lower than 26%, while using the Sliding-Window Ensemble Framework. At the time this paper provided a real novel approach due to using it in a real-time testbed simulation, while other works were mainly tested offline. On the other hand for this model is the recommendation system still in development and this model considers a regular taxi service instead the offer Uber's offer approach, i.e. is not accounting for passengers that can refuse the offer.

In the work [22] is the hotness index proposed, which indicates in what cluster a taxi is needed most. The hotness index is obtained by mining historical data to predict the demand distributions with respect to the contexts of the time, the weather, and the taxi location. The four-step process consists of data filtering, clustering, semantic annotation, and hotness calculation. In this work are mainly examined the effects of three different clustering algorithms. This work introduces a similar approach to the one proposed in this thesis, while rating clusters/regions by its desirability to determine where to send a taxi, to fill the demand. However this work considers the density of the number of the offers in the cluster and its distance to the driver, although it lacks the expected profit from the journey and this point of view is not considered.

Chapter 3

System Model

In this chapter we introduced an agent-based model for the on-demand transport mechanism. In particular, our model accounts for the topology of a city, the profit-based motivations of the drivers, and the preferences (i.e., trip distances and acceptable prices) of the passengers. The request of each passenger is based on desired pick-up and drop-off locations and times, which are communicated to the transport provider that matches the passengers to drivers and prices for the passengers.

In the remainder of this chapter is first detailed the map representation, followed by the modelling details of each agent, e.g passengers, drivers and the provider.

Note that to create this model more clearly some assumptions are introduced, and are notated as capital **A** followed by a number appendix.

3.1 Map representation

The network of roads that drivers can use is modeled via a directed graph $G = (V, E)$. In the graph, the set of nodes V represents possible pick-up and drop-off locations of passengers. The set of edges E represents the direct routes between the locations in V, which can be traversed by the drivers. Associated to each edge $e \in E$ are:

1. a pick-up location $u \in V$;
2. a drop-off location $w \in V$;
3. a cost $c_e \in [0, \infty)$ for a vehicle to traverse edge $e \in E$, due to fuel consumption as well as vehicle wear and tear;
4. and an edge traversal time $\tau_e \in \mathbb{Z}_+$.

3.2 Types of agents

The system consists of three types of agents: N passengers; K drivers; and the mobility broker, provider of the on-demand transport service.

A1: Ridesharing is not supported; that is, each driver can only transport a group of passengers that have the same pick-up and drop-off locations and times.

A2: Passengers enter the network at different times.

3.2.1 Passenger Agents

A new passenger enters the network when it makes a journey request to the dispatcher ,mobility broker, which for passenger i consists of:

1. a pick-up location $v_{i,p} \in V$;
2. and a drop-off location $v_{i,d} \in V$.

A3: The edge cost c_e and edge traversal time τ_e are computed offline during pre-processing where the mobility broker solves the shortest path problem between u and v on the underlying road network.

A4: We assume that the system is operating within a taxi spot market. As such, the passenger requests immediate pick-up and cannot prebook a taxi.

When a passenger requests a journey, the mobility broker makes an offer that the passenger can accept or reject.

A5: We assume that each passenger's decision is based on two factors: the maximum price, $p_{i,\max}$, that the passenger is prepared to pay for the journey; and the maximum deviation, $\Delta \in [0, \infty)$, representing the maximum time the passenger is prepared to wait. These can be viewed as the passengers' types [19]. More precisely, the deviation is defined as follows.

Definition 1. *Let T_i be the actual pick-up time for passenger i and a_i be the time that the passenger made the journey request. The deviation, denoted by δ_i, is defined as*

$$\delta_i = \begin{cases} T_i - a_i, & \text{if } T_i - a_i > 0 \\ 0, & \text{otherwise.} \end{cases} \tag{3.1}$$

Let p_i be the price offered to passenger i for his journey.

A6: Passenger i then uses the following policy to decide whether or not to accept the journey offer.

Accept if $p_i \leq p_{i,\max}$ and $\delta_i \leq \Delta$.
Reject otherwise.

The price offered to passenger i is given by $p_i = r_i \cdot R_i$, where R_i is the distance of the passenger's requested journey and r_i is the price-rate (in euros/km) set by the mobility broker. Although non-linear pricing schemes have been proposed [12, 13], the benefits have only been demonstrated in the dispatcher model, where the price is only a function of R_i or the driver commission-rate is not accounted for. This is not the case in our mechanism, where the price-rate can vary (see Section 3.3). As such, $p_{i,\max} = r_{i,\max} R_i$, where $r_{i,\max}$ is the maximum price-rate that passenger i is prepared to pay.

A7: $r_{i,\max}$ is modelled as a random variable, which is independent and identically distributed (i.i.d) for each passenger i. This captures differences in price-rate preferences between passengers. In particular, we assume that $r_{i,\max}$ is beta distributed, with probability density function

$$f_{r_{i,\max}}(x_r) = \frac{1}{{}_{\max}B(\alpha_r, \beta_r)} \left(\frac{x_r}{\rho_{\max}}\right)^{\alpha_r - 1} \times \left(1 - \frac{x_r}{\rho_{\max}}\right)^{\beta_r - 1} \quad (3.2)$$

where $B(\alpha_r, \beta_r) = \frac{\Gamma(\alpha_r)\Gamma(\beta_r)}{\Gamma(\alpha_r+\beta_r)}$ is the beta function, $r_{i,\max}$ has support $[0, \rho_{\max}]$. This is a flexible model, which can capture a variety of scenarios, such as different times of day simply by varying the parameters α_r and β_r.

A8: We assume that the support $\rho_{\max}$ depends on the location of the requesting passenger, i.e. on the desirability of the region.

$$\rho_{\max} = \rho_{\max} * d; \quad (3.3)$$

Moreover model of price of request $p_{n,t}$ and cost of request model $c_{n,t}$ are introduced for computational purposes. A simple, general model is to treat $p_{n,t}$ and $c_{n,t}$ as Beta distributed random variables. The are three parameters that define these distributions: α; β; and ρ. Each of this parameter is corresponding to each of $p_{n,t}$ and $c_{n,t}$, so totally there are 6 parameters. Importantly, ρ corresponds to the maximum price and cost, respectively. This is reasonable as the area of the serviced region, the maximum price passengers are prepared to pay, and the commission-rate, are all bounded.

A9: In summary it is assumed that the net profit is characterized by two probability density functions $p_{n,t}$ and $c_{n,t}$, which are both beta distributed and is computed as $p_{n,t} - c_{n,t}$.

More precisely, the distributions are defined as follows:

$$\begin{aligned} f_{p_{n,t}}(x) &= \frac{1}{\rho_{p_{n,t}} B(\alpha_{p_{n,t}}, \beta_{p_{n,t}})} \left(\frac{x}{\rho_{p_{n,t}}}\right)^{\alpha_{p_{n,t}}-1} \left(1 - \frac{x}{\rho_{p_{n,t}}}\right)^{\beta_{p_{n,t}}-1} \\ f_{c_{n,t}}(x) &= \frac{1}{\rho_{c_{n,t}} B(\alpha_{c_{n,t}}, \beta_{c_{n,t}})} \left(\frac{x}{\rho_{c_{n,t}}}\right)^{\alpha_{c_{n,t}}-1} \left(1 - \frac{x}{\rho_{c_{n,t}}}\right)^{\beta_{c_{n,t}}-1}, \end{aligned} \quad (3.4)$$

where

$$B(\alpha, \beta) = \frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha + \beta)}, \quad (3.5)$$

and $\Gamma(\cdot)$ is the Gamma function (a standard special function). Note that all parameters (i.e., $\rho_{p_{n,t}}, \alpha_{p_{n,t}}, \beta_{p_{n,t}}, \rho_{c_{n,t}}, \alpha_{c_{n,t}}, \beta_{c_{n,t}}$) are estimated from data at each 1.

3.2.2 Driver Agents

In order to represent the taxi fleet, we need to model agents, that will represent the drivers with the vehicles then being attached to them automatically. Once passenger i accepts his journey, the driver , to which will be the passenger i assigned is chosen randomly from the set of suitable drivers, while drivers that are in the same region are prioritized, in order to reduce the fuel cost, and also in order to properly exploit the drop-off routing feature. Generally the taxi drivers compete against each other to get a customer, so the probability of getting him differs based on their strategies, but for this research this fact not important. This randomness should simulate the auction mechanism used in the original system model [38].

A10: The probability of a taxi to be assigned to new-requesting customer, differs on:

(a) If the taxi is located in the same region as the passenger, the probability of getting it is equal to all other taxis in the region n during the time period t, thus is equal to $\frac{1}{m_{n,t}}$, where $m_{n,t}$ is the number of taxis in the region n in the time period t.

(b) If the taxi is not located in the same region as the passenger, the probability of getting it is equal to all other taxis in that are suitable for serving the customer i.e., fulfill the condition of being able to serve the customer within the maximum deviation Δ 1 during the time period t, then the probability is equal to $\frac{1}{m_{\Delta,t}}$, where $m_{\Delta,t}$ is the number of taxi drivers located within the maximum deviation Δ from the customer.

The drivers pay a commission-rate to the provider. This commission-rate represents the fraction of revenue from the passenger that the driver is willing to pay to the mobility broker. As such, η_j determines driver j's profit for the journey, given by

$$S_{j,i} = (1 - \eta_j) r_i R_i - c_i \tag{3.6}$$

where c_i is the cost of transporting passenger i, given by $c_i = c(R_i + R_{j,i})$ with c denoting the cost-rate (in euros/km) and $R_{j,i}$ denoting the distance from driver j's initial location to passenger i's pick-up location.

A11: In order for driver j to bid for a passenger at commission-rate η_j, the journey profit must satisfy

$$S_{\min,j} \tau_j \leq (1 - \eta_j) r_i R_i - c(R_i + R_{j,i}) \tag{3.7}$$

where $S_{\min,j}$ is the minimum profit per minute the driver will accept, and τ_j is the total duration of the journey (from the driver's initial location to the passenger's drop-off). Importantly, the minimum profit per minute, $S_{\min,j}$, is only known to the driver.

A12: The commission-rate is set by default as $\eta_j = 0.05$ i.e. 5% of the driver's raw profit.

The drivers have an objective to maximize the profit $S_{j,i}$ from each ride, maximize the short term profit which will lead to maximizing their long term profit.

3.2.3 Provider Agent

The function of the provider (mobility broker) agent is to price passengers, and select and pay drivers (detailed in Section 3.3). In order to perform this function, the mobility broker requires statistical knowledge of the maximum price-rate, $r_{i,\max}$, each passenger is prepared to pay.

A13: The statistical knowledge of the maximum price-rate must be learned. This means that the distribution that the mobility broker obtains will in general not exactly correspond to the actual distribution of $r_{i,\max}$, given by (3.2). As such, the probability density function $\hat{f}$ known to the mobility broker is modeled by the Beta distribution; i.e., the same as (3.2), except with parameters α'_r, β'_r. The parameters α'_r, β'_r are typically not equal to the parameters α_r, β_r.

The personal preferences of the provider is to maximize the fraction of profit he gets from each ride. More precisely its represented with the following equation:

$$\begin{aligned} \underset{r}{\text{maximize}} \quad & rR_i \int_r^{\rho_{\max}} \eta \cdot \hat{f}(r)dr \\ \text{subject to} \quad & 0 \leq r \leq \rho_{\max} \end{aligned} \tag{3.8}$$

where

$$\hat{f}(r) = \frac{1}{\rho_{\max} B(\alpha'_r, \beta'_r)} \left(\frac{r}{\rho_{\max}}\right)^{\alpha'_r - 1} \left(1 - \frac{r}{\rho_{\max}}\right)^{\beta'_r - 1} \tag{3.9}$$

3.3 The Mechanism

Now the mechanism, that will allocate and price passengers as well as drivers, is being introduced. The final aspect of the model is the mechanism used by the provider to match passengers to drivers, price passengers and determine driver payments. The prices are optimized by the provider to account for knowledge of the probability density functions for each passenger's maximum price rate and each driver's minimum profit. The mechanism forms a two-sided market where prices are set using posted prices [5] (as opposed to an auction-based approach), which is inspired by Uber's approach.

Passenger prices are obtained by solving the following optimization problem at each location $k \in \mathcal{L}$

$$\underset{r}{\text{maximize}} \quad r \int_r^{\rho_{r,k,\max}} f_{r_{i,k,\max}}(r)dr. \tag{3.10}$$

Intuitively, this optimization problem corresponds to the expected revenue from each passenger. The provider aims to maximize the revenue for the passengers. Note that competition between passengers is implicit in our formulation; that is, the parameters $\alpha_{r,k}, \beta_{r,k}, \rho_{r,k,\max}$ capture this competition. For example, in regions with higher competition, $\rho_{r,k,\max}$ (the support of the distribution), is typically larger.

On the other hand, the driver payments are obtained by solving

$$\begin{aligned} \underset{\eta_{\min}}{\text{minimize}} \quad & \eta_{\min} \\ \text{subject to} \quad & \int_{\eta_{\min}}^{\rho_{s,\max}} g_{\eta_{j,\min}}(\eta_{\min}) d\eta_{\min} \geq 1 - \epsilon \end{aligned} \tag{3.11}$$

where ϵ is the minimum probability that each driver will not accept the offer.

Intuitively, the optimization problem in (3.11) selects the commission rate $\eta_{\min}$ to ensure the probability that a driver will accept is sufficiently high, while also ensuring that the service is profitable for the provider.

Chapter 4

Proposed Approach

In this Chapter two recommendation systems are proposed. First, the mathematical formulation of the problem is developed, which provides a simplified mode; of the on-demand transport service system. This simplified model of the system is used in the recommendation systems, making the routing problem tractable. This is achieved by using a simple parametric model.

Later on the two proposed approaches (*profit-aware* and *customer-aware*) are introduced. The aim of both approaches is to exploit the information about the system in order to provide an efficient solution for the routing problem through the recommendation system.

Recommendation system is a program that routes the drivers towards the next location, the word recommendation is crucial because the drivers are self-interested and they are not obliged to follow the directions of this system if it is against their intentions.

4.0.0.1 Time discretization

Due to the time variability of the requests, following simplification of the model is implemented. The continuous time is discretized by τ minutes e.g. 30 minutes, this means that the time interval $[t, t+1)$ is equal to a 30 minutes, similar as used in the work [9]. The solution for alternating crowdedness of a region, is more precisely suggested in the paper [24], but because the computation complexity is also crucial for this approach, it is used less precise, but computational faster discrete time intervals. Note that the passengers still enter the network at different times according to the continuous approach. This discretization is only used for the implementation purposes, and as a model of passengers behaviour it appears as sufficient.

A11: It is assumed that passengers arrive according to a homogeneous Poisson process $\Pr(\mathcal{P}_{n,t})$ with parameter $\lambda_{n,t}$ i.e., the mean number of passengers arriving in the interval $[t, t+\tau)$ at 1 n is constant.

Definition 2. *A key theorem for Poisson processes is the distribution of the number of arrivals in a given time interval, which is stated as follows.*

$$\Pr(N(t+\tau) - N(t) = k) = \frac{e^{-\lambda_{n,t}\tau}(\lambda_{n,t}\tau)^k}{k!}. \tag{4.1}$$

Note that the distribution is known as the Poisson distribution. A Poisson process is defined on page 247 in Grimmett and Stirzaker [21].

4.0.1 Probabilistic model of the system

The main difference between the probabilistic model used for the recommendation system and the system model, is that it exploits simple parametric models to describe complex models of the system.

In particular beta distribution R_i is used to describe the distance of each journey, which is later used to calculate the expected profit and also cost from each fare. To calculate the number of arrivals in certain region a Poisson distribution is exploited.

Later in this Chapter more variables needs to be calculated, namely:

1. $\mathbb{E}[p_{n,t} - c_{n,t}]$ describes the expected value of the net income from a customer. This value depends on the raw profit $p_{n,t}$, that is computed as the the price rate r_i (differs on the type of the region) multiplied by the probable distance $R_{n,t}$ of the trip and the cost of the trip $c_{n,t}$, k_c denotes the cost per kilometer k_c [EUR/km].

   Theorem 1. *The price per kilometer r_i is clearly dependant on the region n, see section 4.0.2 and represents the average price per kilometer obtained from the data. The cost per kilometer k_c is a chosen value* 0.3 *[EUR/km], that depends on the used cars and gas prices.*

$$\mathbb{E}[p_{n,t} - c_{n,t}] = \mathbb{E}[r_i * R_{n,t} - k_c * R_{n,t}] \tag{4.2}$$

2. The expected profit for the driver depends exponentially on $\Pr(\mathcal{P}_{n,t})$ i.e how many customers are appearing in the current location and also how many taxis are competing against this taxi driver in order to serve this customer. This clearly depends on the number of available taxis in the region $m_{n,t}$ and how long the driver waits for a customer in the region and the probability $\Pr(\mathcal{A}_{n,t})$ that the customer will accept the proposed price for the journey, in region n during time period t, and it is obtained from the previous data. The probability distribution describing the number of passengers arriving according to a homogeneous Poisson process with parameter $\lambda_{n,t}$ in the time interval $[t, t+\tau)$, for which the driver is selected (with probability $1/m_{n,t}$). To compute the probability we need to thin the Poisson process 2 so that only the passengers for which the driver is selected are considered.

   Theorem 2. *This theorem holds if only the taxis that are in the same region can serve the customers (the m_t variable represents the reality truthfully). Note that instead of continuous time are used time intervals τ with length 30minutes*

$$\Pr(\mathcal{P}_{n,t}) = \Pr(N(t+\tau) - N(t) > 0) = 1 - e^{-\frac{\lambda_{n,t} \cdot \tau \cdot \Pr(\mathcal{A}_{n,t})}{m_{n,t}}} \tag{4.3}$$

3. Accounting for the cost $c_{0,n}$ represents the intention to lower the gas and CO2 consumption and also the drivers intention to maximize its profit, which is greatly affected by the driven distance (cost).

Theorem 3. *The cost $c_{0,n}$ depends on the cost per kilometer k_c [EUR/km] and $R_{0,n}$ [km], the distance from taxi's current location to the region n.*

$$c_{0,n} = k_c \cdot R_{o,n}. \tag{4.4}$$

4. η is chosen for the whole system identically and its value is 5%.

4.0.2 Regions

Based on the character system model graph representation, which includes a high amount of nodes, the concept of *regions* is introduced. Region is defined as a subset of nodes V that lie withing a certain area. The intention behind this idea is to lower the computational complexity and also it is taken into account that taxis are able to serve customers not only at the actual place, where they are waiting but also within a certain radius. Each node is assigned to a region depending on their latitude and longitude. The map division has a grid character and each region is a $1km^2$ big square cell of this grid.

Note that the concept of the regions is only used for the computational purposes, the underlying graph based network is untouched, so that driver and passengers still travel through nodes. This feature preserves the precision of computation and also improves the computation, in terms of spread and also better representation of the proposed approach.

The improvement is significant also in terms of accounting nearby nodes, that lie within the driving distance, where can the driver serve the customer, into the computation. The regions are equally big in terms of covered area, to allow the taxi driver to reach each part of it in reasonable time. This approach was also adopted in [25, 23, 14]. The idea of the grid is used instead of the well formed clusters used in [23], because grid approach is completely sufficient and clusters are not quite suited for our purpose . The clusters have different size and this brings an uncertainty into the potential computation, since there are a various travel costs from the center of the cluster to the customer, that might be located on the edge. This means that the cost to get to the customer from the centre, where the taxi is sent, may vary for each region, while using the grid concept assures that the distance is the same for every cell.

Note that in the experiments chapter 6 are tested scenarios when the ρ variable in the pricing formula 3.2 differs from region to region. This should illustrate that some places across the city has larger desireability or different type of customers, i.e. customers that are prepared to pay more, than in other places.

4.0.3 Profit-aware recommendation approach

In order to provide a recommendation system for the novel smart companies such as Uber [34], and to provide a better customer oriented service, more information needs to be taken into account than used in the *customer-aware* approach once modelled and implemented a *profit-aware* approach that considers more information about the customers.

In this proposed approach, locations are suggested based on the expected profit a driver will receive on his next journey from each location. As such, the approach is able to capture passenger price preferences, the distance of the next journey, the passenger arrival rate, the number of competing drivers at each location, as well as the driver's travel cost. Such preferences have not been considered in previous approaches to the driver placement problem, all of which are targeted at traditional taxi services. For the computation a parametric statistical model is exploited, that has a small number of parameters to estimate and has the desirable feature that the expected profit for the next trip is computed in closed form. [32].

The profit-aware approach routes the taxi driver towards the expected customer, while taking into account his profitability. Importantly, the number of passengers likely to be arriving at each location, the price the passengers are likely to pay, the number and location of other taxis, expenses that are associated with routing and of course the time dependence, are all accounted for. So that the taxi fleet is distributed across the city in respect to the expected demand, to cover more customer requests. Essentially the intention is not to serve as many customers as possible but to seek the most profitable ones. This will increase the overall revenue of the taxi driver, as well as the profit of the provider.

The goal of our algorithm is to send drivers without a passenger to the location where the expected profit the driver will receive for the next trip is maximized. Our algorithm selects the hotspot (most profitable location) from a set of regions, to route the driver towards it after the customer drop-off. Importantly the cost to the hotspot is considered in the potential because the taxi driver is going to use fuel and wear out the car. This property ensures that taxi drivers will be more likely stay in the close regions in order to lower the fuel consumption and related emissions. Hotspot location is selected based on a potential function, which ranks the hotspots based on the expected profit a driver will receive on her next journey. The potential function takes also into account the space-time variability of the system. Meaning that the potential is different for each region and also for each time period. This dependence provides a higher prediction precision.

Note that in this approach when the expected profit from a possible routed trip is negative, then this routing is not realized, because the expected profit that would lead this routing towards a possible customer yield is negative for the driver, which respond to driver's personal preferences.

4.0.3.1 Potential Function

The potential represents the expected net income that the taxi driver is likely to earn in the region n during time period t. The potential is computed as the net profit from the expected request that will be assigned to the driver. The formula is following:

$$o_{n,t} = \mathbb{E}[p_{n,t} - c_{n,t} | \mathcal{P}_{n,t}, \mathcal{A}_{n,t}] \Pr(\mathcal{P}_{n,t}, \mathcal{A}_{n,t}) - c_{0,n} \tag{4.5}$$

The potential value depends directly on the expected profit $\mathbb{E}[p_{n,t} - c_{n,t}]$, which is a random variable, that is multiplied by the probability $\Pr(\mathcal{P}_{n,t})$ that this request will be assigned to the driver. It also takes into account the cost $c_{0,n}$ of driving to the region n, which is known exactly. All the variables are computed in the Section 4.0.1.

4.0.4 Customer-aware recommendation approach

Based on the previous work [14, 22, 9, 10], where the goal is to find the supply demand equilibrium i.e. successfully cover the demand, it is proposed a customer-aware approach, which the objective to serve as many customers as possible, by taking into account the overall driven distance, that is reflecting the fuel consumption and connected CO2 emissions, by the taxi fleet and also increase the amount of time of each taxi is being *occupied*. This approach uses simplified heuristic algorithm that accounts for the number of incoming requests, number of taxis over the map and travelled distance, but does not account for the price preferences of passengers. Note that in comparison to the other approaches that often use ARIMA algorithm for recommendation, this approach provides a easy-to-implement and elegant way of doing so.

This algorithm uses a spatiotemporal potential, that is computed from the mentioned variables, to rate the locations on the map based on their desirability, it is closely described in the next Section.

4.0.4.1 Potential Function

This approach does not account for the expected profit of each customer, rather for customer i.e. request itself. Note that it also can be used in the classical markets, while it does not consider the probability that passengers might reject the journey. So that of the concept and structure of the algorithm remains the same as 1 except that the potential function. The function is adjusted accordingly:

$$o_{n,t} = -c_{0,n} * \frac{P_{n,t}}{m_{n,t}} \tag{4.6}$$

Where $c_{0,n}$ is the distance to the region n, and $\frac{P_{n,t}}{m_{n,t}}$ represents the ratio between requesting passengers and the number of taxis in the current region that are able to serve them. The potential now represents the cost of hunting an expected passenger, while considering the demand. This approach represents the pursuit of covering the demand, regardless of their profitability, while minimizing the cost of routing.

Note that the possibility that passengers can refuse the offer is not considered.

Note that this is not a profit-oriented approach, and also the character of the customer-aware potential allows it to be negative while in the profit-aware approach a trip is not routed when the expected profit is negative.

Chapter 5

Implementation

In this Chapter the implementation of the recommendation system is discussed. The recommendation system is incorporated into the Mobility testbed simulation which is coded in Java language. Mobility Testbed is an interaction-rich simulation tool for testing and evaluating control mechanisms for traditional demand-responsive transport services (Dynamic Dial-A-Ride Problem) as well as next-generation flexible mobility services (exmplified e.g. by Uber or Lyft) [38].

Despite the novel approaches and precise models of previous works, most of them are tested on offline testbeds, nevertheless the work [9] has introduced a model that works in real-time simulation with time window of 30minutes, however the recommendation system is still not proposed.

The base of this this was to implement and experimentally evaluate the acquisition that could bring a profit-aware recommendation approach for the new smart companies. Based on this intention there were proposed two approaches mentioned before, in order to properly evaluate the novel profit-aware approach. Moreover the used model takes into account the personal preferences, not known to the provider, of all three participants of the service, the provider, the driver and the customer. Crucially, this approach explicitly accounts for private passenger preferences, not known to the provider. In particular it is not assumed that all passengers will accept a journey at the market price.

To evaluate both driver placement algorithms, a simulation study is performed of an on-demand transport service based in the Hague. We consider three perspectives: passengers; drivers; and the provider. Our simulation results show that accounting for whether or not a passenger accepts an offer at market price matters, particularly when there is a variation in demand over a city. We demonstrate this fact by comparing our algorithm for on-demand transport markets with our simplified heuristic approach and a baseline approach based based on the traditional taxi model. We show that our algorithm for transport markets can outperform both alternative approaches in terms of total revenue and driver occupancy [32].

Since this work does not operate with an existing taxi fleet, but is using a real-world test bed simulation, it is able to change the number of taxis or requests, variously, which provides a good tool to evaluate various scenarios and settings of the input parameters, the experiments provide a useful insight for the provider. In other words they provide a good evaluation and also suggestion for the provider on what he should be focusing in order to improve a less efficient part of his service.

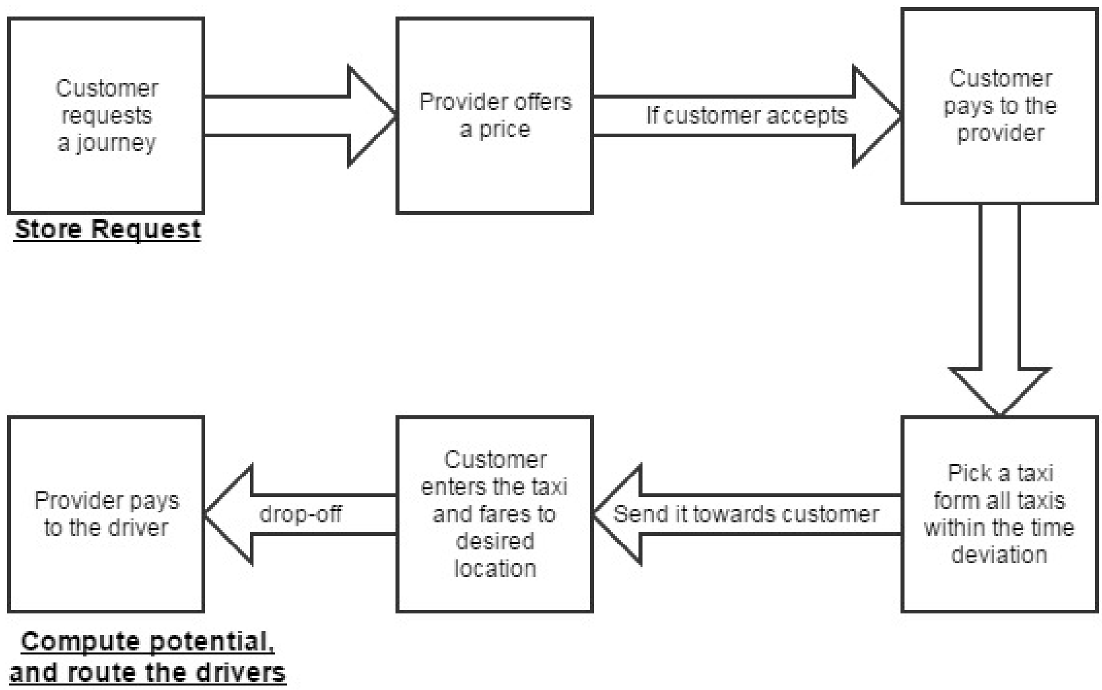

Figure 5.1: Routine diagram

The main routine of the program, that describes the market mechanism is described in the following diagram 5.1. The bold underlined text indicates the part of the routine that has been changed in order to implement the recommendation system. Except for two major changes, the computation of the mathematical model is being run during the initialization.

In the remainder the part of the code containing the recommendation system is discussed.

5.1 Dispatching Logic routine

The task of the proposed algorithm is to estimate the location of next profitable customer from the previous data. The algorithm implements following functions that are necessary for running the routing feature:

- Storing the data from previous requests
- Computing probabilistic models from previous data
- Evaluating the potential of the nodes
- Routing taxi drivers after drop off

As mentioned before this algorithm is implemented into an existing agent-based model framework [33], which provides all other necessary methods and classes such as finding the path among nodes, computing the distances, generating scenarios etc.

The algorithm is implemented into the *MyDispatchingLogic.java* class, which simplified structure is described in the algorithm 1. This class puts all the other classes together and puts them to desired use. Namely they are:

- **Saver.java**: stores requests in desired format
- **Miner.java**: merges old and new data together.
- **MyProbabilityMath.java**: uses mined data to compute the distributions poisson and beta distributions. Computes the potential.
- **Regions.java**: divides the map into regions, and computes to which region each position belong, based on longitude and lattitude.
- **Stats.java**: handles the stats of drivers, and provider, i.e. computes the profit, time and distance driven.

The *MyDispatchingLogic* class it consists out of three main methods *initializeModels()*, *newRequest()* and *dropOff()*.

5.1.1 Initialization

The first, initializing method loads the previous data from files where they have been previously stored and computes the distributions for each region n in each time interval t_i of:

- Poisson distribution for number of passengers arriving
- Beta distribution for fare prices
- Beta distribution for fare costs

These distribution are needed for computing the potential of nodes and thus routing the taxis towards profitable customers.

5.1.2 Processing an incoming request

The processing of new request consists of following steps:

- Store current request to file
- Find taxis that are eligible to complying with request
- Pick taxi and send him to serve the request

Then the serving taxi is picked randomly from a set of suitable ones, which should simulate the auction of drivers. After the taxi is picked the taxi driver is sent via the shortest path (computed during pre-processing) towards customers location, and then again via the shortest path to the customer's desired location. The method $passengerWillingToPay()$ corresponds to the Equation 3.10.

5.1.3 Finishing the request

Once the requested journey is accomplished, the customer is *dropped off*, and the then the potential for each node n in current time period t_i is computed. The node with maximal potential $o_{n,t}$ is set as next destination for the driver, where he should meet the next customer.

```
Data: Scenario
Result: Fares assigned to taxi drivers
begin initializeModels
    computeRegions();
    loadDistributionFunctions();
    computeAcceptanceRate();
    initializeDataFiles();
    initializeFieldsAndPositions();
end
begin newRequest
    price = getPricePerKM() * request.getDistance() ;
    if price < passengerWillingToPay then
        foreach taxi in FreeTaxis do
            if taxi.isEligible() then
                eligibleTaxis.put(taxi);
            end
        end
        servingTaxi = pickRandomTaxiFrom(eligibleTaxis);
        servingTaxi.assignRequest(request, price);
        servingTaxi.recievePayment(price);
    end
end
begin dropOff
    maxPotentialRegion = ;
    computeMaxPotential(Regions, currentTime, positionOfFreeTaxis);
    taxi.sendTo(maxPotentialRegion);
end
```

Algorithm 1: Dispatching Logic

5.1.4 Implementation-based model adjustments

In order to be able to study a larger span of data simulations following adjustments are introduced. The historical data of requests are thinned, so that only requests, that are accepted by the customers, are considered. This does not influence the performance of the recommendation systems. Since this fact is not incorporated into the Mobility Testbed simulation, following adjustment are used.

In the algorithm 1 the customer always accepts the proposed price. This means that the set of refused requests is not independent, and these requests have a common characteristics. Neverless the acceptance rate is computed and accounted for in the poisson distribution of the customer arrival rate.

Note that the customer-aware approach takes into account all customers and does not count with the possibility of refusing a request, and needs to be adjusted following:

$$o_{n,t} = -c_{0,n} * \frac{P_{n,t}}{m_{n,t} \cdot \Pr(\mathcal{A}_{n,t})}$$

In the profit-aware the only change the probability that passenger accepts the price $\Pr(\mathcal{A}_{n,t})$ to 1 i.e. 100% (it does not affect the computation), because from the learned data only the accepted journeys are considered.

The random approach remains unaffected since the change does not affects it.

Chapter 6

Experiments

In this section we perform experiments to evaluate the recommendation system.

6.0.5 Measured performance criteria

To provide relevant comparisons of our proposed *Profit-aware routing* recommendation system, it is compared alone *No routing* system, does not use the drop-off routing feature, and also with the *Customer-aware routing* recommendation system. The *Customer-aware routing* approach is in the same spirit as other approaches targeted at traditional taxi models, with the important difference that the heuristic is based in part on the same parametric statistical model used for our approach targeted at services exploiting a market mechanism. This means that parameters for the heuristic are simple to estimate and the heuristic can be computed in closed form. All approaches are coded into the model, and are using the same evaluation methods and data, and scenarios, which ensure an objective evaluation [32].

There are 5 different optimization criteria that are being examined:

1. Total revenue of the provider - represents the personal preference of the provider.
2. Total driven distance of all drivers - represents the general preference of lowering the consumption of gas and CO2 emissions.
3. Average profit per driver - represents the personal preference of the drivers.
4. Average *occupied* time - represents the personal preference of the drivers.
5. Ratio of successful requests - represents the personal preference of the customer, how many requests can be served by using different approaches in distributing the taxi fleet over the map.

Next figures displays the impact on the performance of low density setting, e.g. 60 request and 2 drivers per 24 hours. The difference between the 3 approaches, when the demand is decreasing is shown.

In each figure, we observe the effect of passenger demand, encapsulated in β (from (3.2), where $\alpha = 1$), which determines also the price-rate each passenger is charged via (3.10). The small β corresponds to high demand and the large β corresponds to low demand. The Figure 6.1 that for small β the probability that customer will accept higher price is bigger, and for large β the customers are more likely to refuse those high price offers, which results to smaller number of accepted journeys and also smaller average price of these journeys.

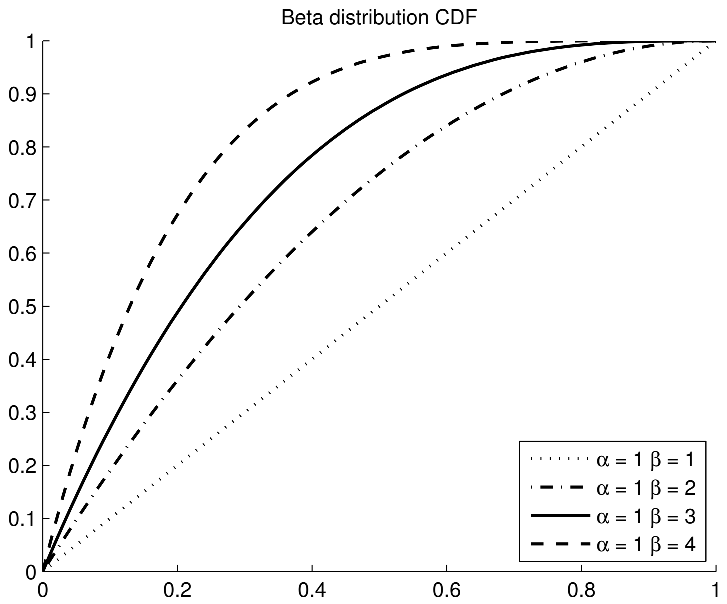

Figure 6.1: CDF for the beta distribution for various β

6.1 Illustrative example

To show the different performance between the *profit-aware* and the *customer-aware* approach during the high demand, an illustrative example of 30 requests and 2 drivers operating is discussed. The different attributes of these approaches indicates that the *profit-aware* As expected the performance is similar when the demand is low.

The performance is examined more precisely, in the next paragraph.

The Figure 6.2 shows the effect of passenger demand on the number of the requests that are accepted. For this performance metric it can be clearly seen that the system without the routing feature, is able to serve only fraction of the customers served by the proposed approaches. This is caused by the fact that the drivers are not close enough to an upcoming request, thus cant serve the customer withing the given time Δ.

The Figure 6.3 shows the effect of passenger demand on the daily provider revenue. Observe that *profit-aware* placement algorithm for on-demand transport markets can outperform the *customer-aware* during the high demand by a long shot.

The Figure 6.4 displays the effect of passenger demand on the average daily driver revenue. This clearly depends on the number of customers served and the operational expenses. Note that *profit-aware* outperforms the *customer-aware* during high demand and is almost equal during the low demand.

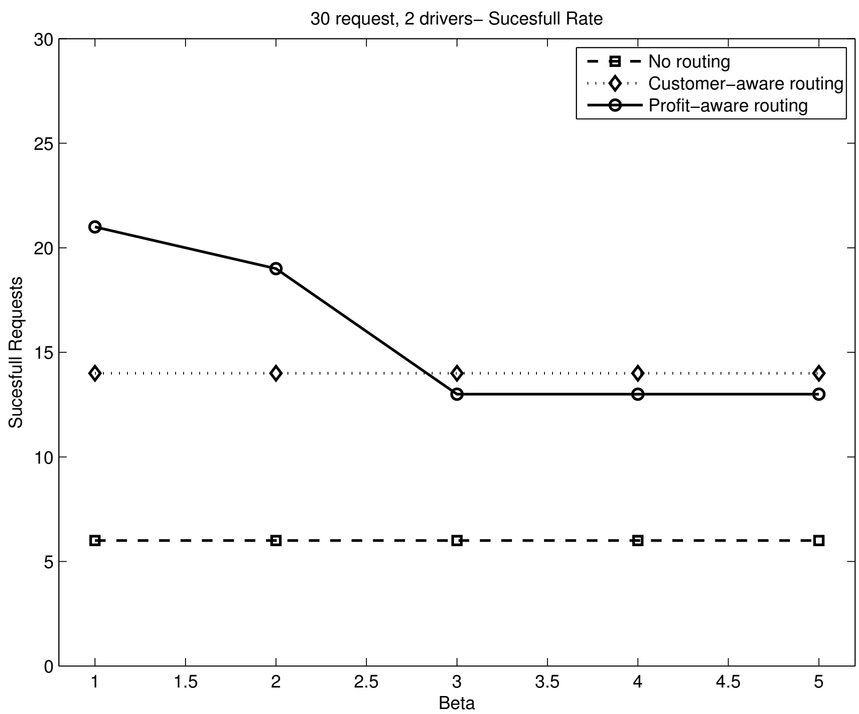

Figure 6.2: Illustrative example: Successful requests

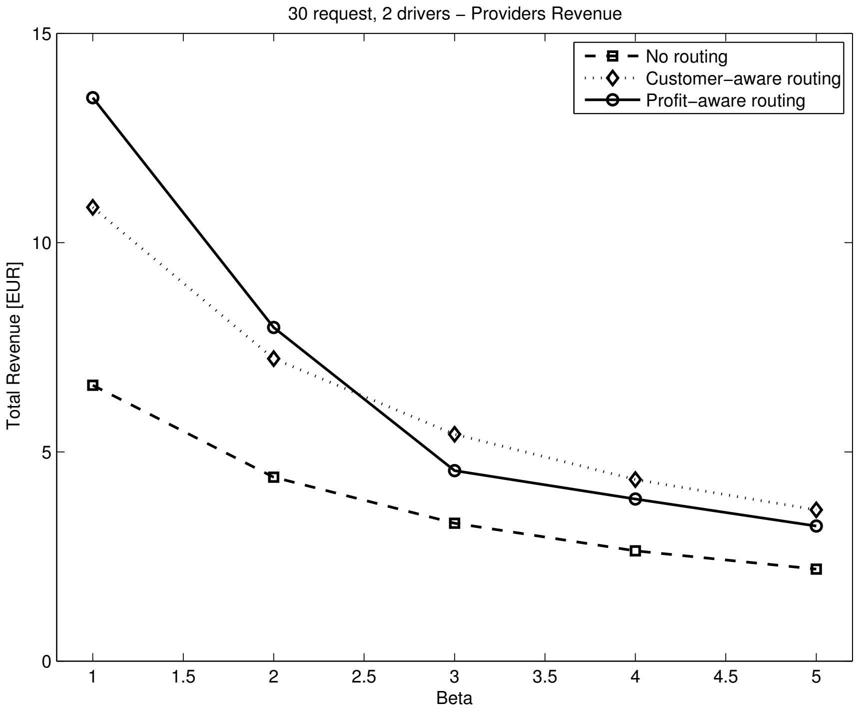

Figure 6.3: Illustrative example: Total revenue

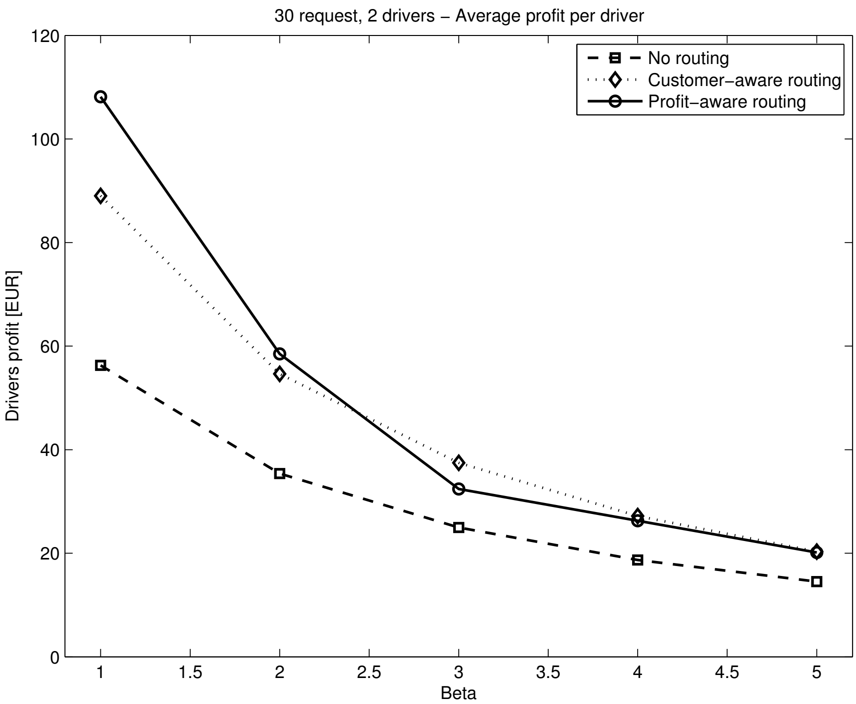

Figure 6.4: Illustrative example: Average driver's profit

6.2 Large-scale experiment

The experiments are run on a map of underlying transport network of the existing city the Hague. The computer setting can be neglected because the routing itself consumes only a little fraction of the computation time, thus comparing the computational times of the different approaches is irrelevant.

The graphs are displayed as comparison of the three approaches by varying the density of operating taxis. The number time window of one run is 24 hours and 300 request occurs during this time period. Each request is ordered 10 minutes before the pickup, which represents the time deviation Δ 1, so that only drivers that are within this deviation can serve the customer. There are 20 iterations of the simulations (corresponding to 20 days), and their average performance is displayed. During each day (one iteration) an average of 600 passengers make a request.

The Figure 6.5 shows the effect of taxi fleet size on the number of the requests that are accepted. For this performance metric it can be clearly seen that the system without the routing feature, is able to serve only fraction of the customers served by the proposed approaches. This is caused by the fact that the drivers are not close enough to an upcoming request, thus can not serve the customer withing the given time Δ.

The Figure 6.6 displays the effect of taxi fleet size on the daily provider revenue. The providers revenue depends only on the commission rate η, which is constant, the price and the number of requests served. The *profit-aware* approach should route the drivers towards more profitable customers this implies that the provider revenue should be bigger when the number of successful requests is less or equal to the *customer-aware* recommendation system. This scenario is shown for the number of taxis is 17 and 23.

The Figure 6.7 shows the effect of taxi fleet size and the driven distance. The driven distance consists of the total distance of all fares and the routing. Observe that the distance is lowest for the *No routing* approach, because there is no routing at all. As expected the *profit-aware* algorithm distance of routing is the highest, because it does not primary prioritize the closest expected customer rather the more profitable one which can be further.

The Figure 6.8 shows the effect of taxi fleet size on the daily driver's profit. Obviously the average driver's profit depends on the raw profit from the fares, which fraction (5%) is displayed in Figure 6.6, minus the operational expenses related to the driven distance per driver. Note that despite almost twice as much driven distance e.g. cost, the driver's are still able to earn more money during the high demand.

The Figure 6.9 displays the effect of taxi fleet size on the average time drivers are *occupied*. This clearly depends on the number and length of the served journeys, and this figure more or less copies the number of successful requests.

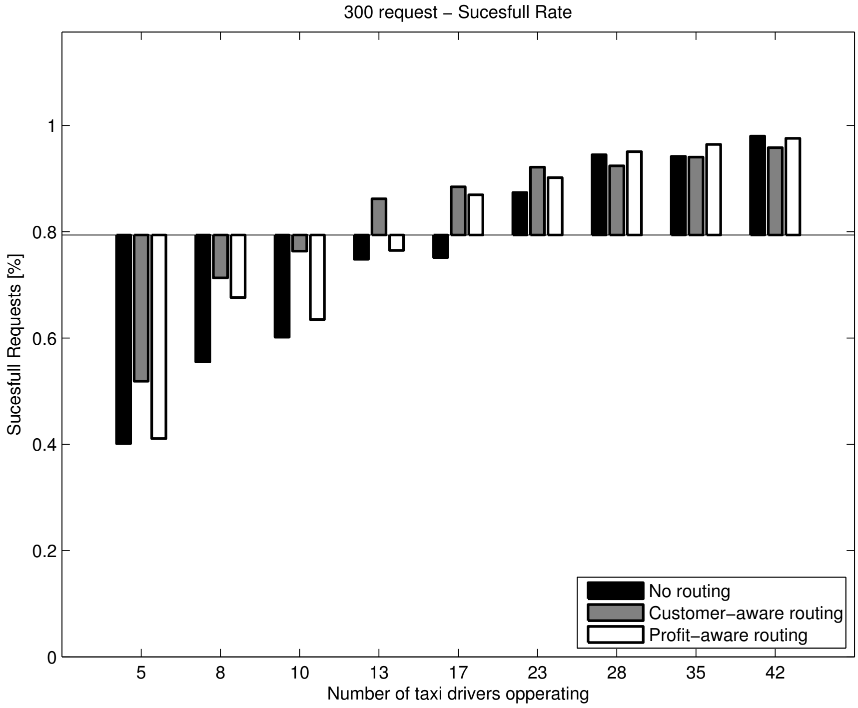

Figure 6.5: Large-scale experiment: Successful requests

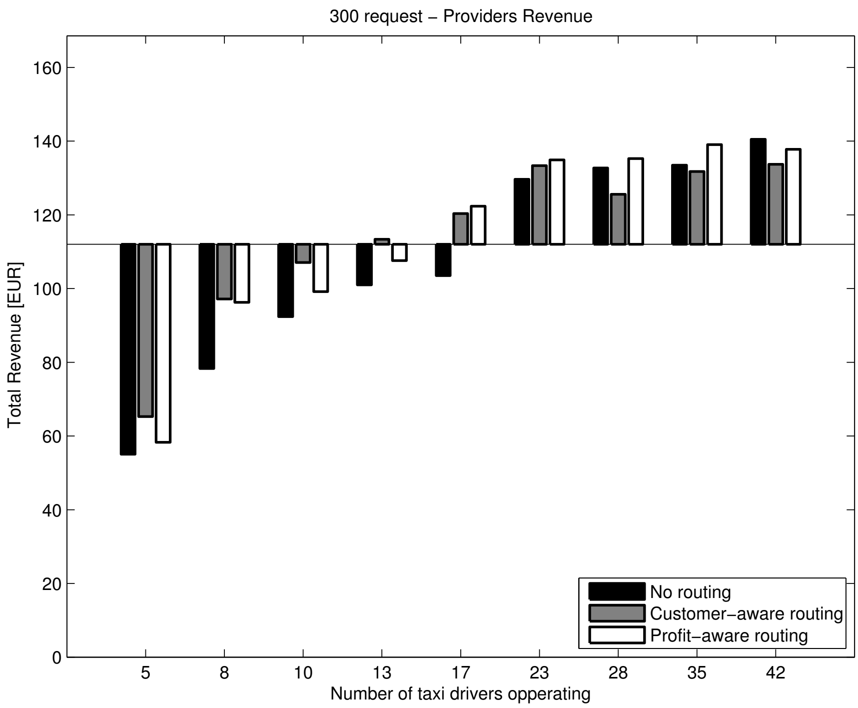

Figure 6.6: Large-scale experiment: Total revenue

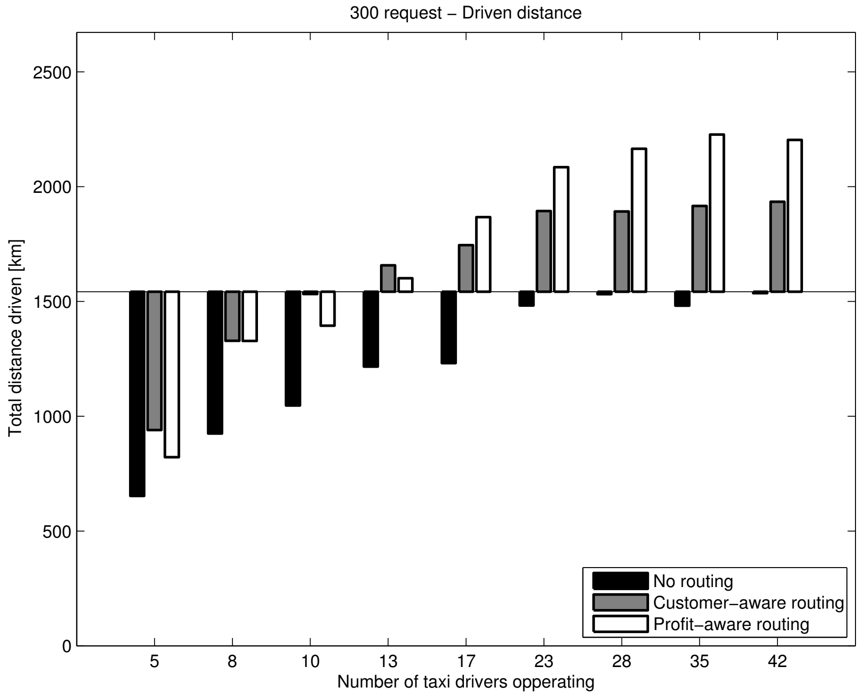

Figure 6.7: Large-scale experiment: Total distance driven

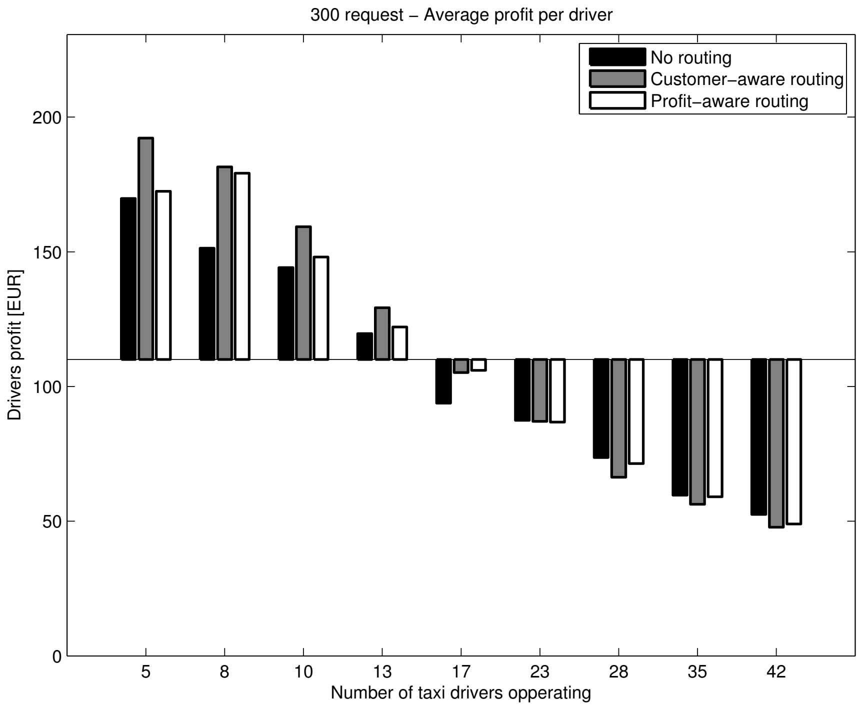

Figure 6.8: Large-scale experiment: Average driver's profit

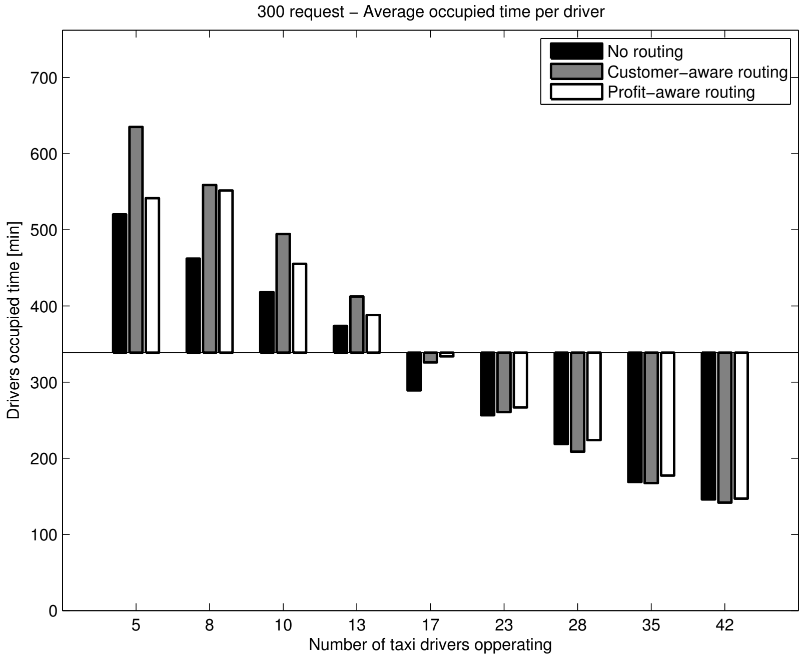

Figure 6.9: Large-scale experiment: Average driver's occupation time

6.2.1 Summary

Generally speaking the number of successful request e.g. the acceptance rate, highly influence all other metrics. Thus in order to gain higher profit and also satisfy the customer's demand which is one of the objectives, it is needed to pay close attention on which approach to choose for which scenario.

1. The *customer-aware* recommendation system proves its quality when the size of taxi fleet is small.

2. On the other hand *profit-aware* has slightly better performance when the size of the fleet is higher.

3. Using a taxi system without the recommendation algorithm for routing, is a good choice when the density of taxis serving is overwhelming, in order to lower the cost of routing, which also increase the profit of the drivers.

Chapter 7

Conclusion

The objective of this thesis is to solve the after drop-off routing problem through modeling, designing algorithms for, implementing, and evaluating the routing recommendation systems.

First the agent-based system model of an on-demand transport service market mechanism, inspired by Uber, is introduced. The models, that represent the personal interests of the agents (customer agents, driver agents, provider agent), are introduced.

To solve the routing problem a mathematical formulation of the problem is created, which provides a base ground for the two recommendation systems are implemented and evaluated. Based on the previous researches a *customer-aware* recommendation system is proposed, which is targeted for the traditional on-demand transport services. Second, novel *profit-aware* recommendation system is designed, that targets for the market-based on-demand transport services, and takes into account more parameters than the previous approach.

Later on a large-scale experiment in an on-line simulation is performed. This experiment studies the impact of the size of the taxi fleet on the general performance of the on-demand transport service, that is using either *customer-aware*, *profit-aware* recommendation system or is not using any.

The experiment has shown that the number of successful request highly influence all other metrics. Which means that trying to get as much customers as possible is the ultimate goal. From the experiment is obvious, that using a recommendation system for the after drop-off routing in the on-demand transport service, represents a huge improvement for both the provider, drivers and customers. The performance of both recommendation systems is depending on the size of the operating taxi fleet and the passenger demand. The *customer-aware* recommendation system performs better when the number of operating taxis is low in comparision to the *profit-aware* recommendation system, that yealds better results when the taxi fleet is bigger.

In the remainder the future work that can lead to better performance is discussed.

7.1 Future Work

This thesis provides a base ground for further research and improvement of the proposed recommendation systems.

1. A more precise model of customers demand can be considered, that accounts for more parameters, not only time, e.g. the weather, the temperature.

2. The recommendation system should route the driver through the low traffic-regions to speed up the journey time and also to avoid unnecessary traffic jams. To do this the traffic should be monitored on-line and this information incorporated into the system.

3. The potential approach should in addition to the potential of the final destination also take into account the potential of the chosen road, because also during routing the taxi is able to serve the customers along the way.

Bibliography

[1] Shoham, Yoav; Leyton-Brown, Kevin (2008). Cambridge University Press. p. 496. ISBN 978-0-521-89943-7 Multiagent Systems: Algorithmic, Game-Theoretic, and Logical Foundations <http://www.masfoundations.org> Multiagent Systems: Algorithmic, Game-Theoretic, and Logical Foundations

[2] E. Kamar and E. Horvitz. Col- laboration and shared plans in the open world: studies of ridesharing. In Proceedings of the 21st International Joint Conference on Artificial Intelligence (IJCAI), 2009.

[3] Cordeau, J. F., Laporte, G. (2007). The dial-a-ride problem: models and algorithms. Annals of Operations Research, 153(1), 29-46.

[4] Borgs, C., Candogan, O., Chayes, J., Lobel, I., Nazerzadeh, H. (2014). Optimal multi-period pricing with service guarantees. Management Science, 60(7), 1792-1811.

[5] S. Chawla, J. Hartline, D. Malec, and B. Sivan, *Multi-parameter mechanism design and sequential posted pricing*, Proc. ACM Conference on Theory of Computing, 2010.

[6] Egan, M., Oren, N., and Jakob, M., Optimizing Profit and Efficiency in On-Demand Transport Services, *Submitted to AAMAS 2015.*

[7] Iosifidis, George, and Iordanis Koutsopoulos. "Double auction mechanisms for resource allocation in autonomous networks." Selected Areas in Communications, IEEE Journal on 28.1 (2010): 95-102.

[8] Yuan, J., Zheng, Y., Zhang, L., Xie, X., and Sun, G., Where to find my next passenger?, *Proc. UbiComp'11* (2011).

[9] Moreira-Matias, L., Gama, J., Ferreira, M., Mendes-Moreira, J., and Damas, L., Predicting taxi-passenger demand using streaming data, *IEEE Transactions on Intelligent Transport Systems* **14**(3) (2013).

[10] Li, B., Zhang, D., Sun, L., Chen, C., Li, S., Qi, G., and Yang, Q., Hunting or waiting? Discovering passenger-finding strategies from a large-scale real-world taxi dataset, *Proc. IEEE International Conference on Managing Ubiquitous Communications and Services* (2011).

[11] Certicky, M., Jakob, M., Pibil, R., and Moler, Z., Agent-based simulation testbed for on-demand transport services (demonstration), *Proc. International Conference on Autonomous Agents and Multi-Agent Systems (AAMAS)* (2014).

[12] H. Yang, C. Fung, K. Wong, and S. Wong. Nonlinear pricing of taxi services. Transportation Research Part A: Policy and Practice, 44(5):337-348, 2010

[13] C. Zeng and N. Oren., *Dynamic taxi pricing* In Proc. European Conference on Artificial Intelligence, 2014

[14] Yuan, N.J., Zheng, Y., Zhang, L., and Xie, X., T-finder: a recommender system for finding passengers and vacant taxis, *IEEE Transactions on Knowledge and Data Engineering*, 2013

[15] Li, X., Pan, G., Wu, Z., Qi, G., Li, S., Zhang, D., Zhang, W., and Wang, Z., Prediction of urban human mobility using large-scale taxi traces and its applications, *Front. Comput. Sci.* **6**(1) (2012).

[16] Horvitz, E., and Krumm, J. Some help on the way: opportunistic routing under uncertainty, *Proc. UbiComp'12* (2012).

[17] Ferreira, N., Poco, J., Vo, H.T., Freire, J., and Silva, C.T., Visual exploration of big spatio-temporal urban data: a study of New York City taxi trips, *IEEE Transactions on Visualization and Computer Graphics* **19**(12) (2013).

[18] Zhang, D., He, T., Lin, S., Munir, S., and Stankovic, J.A., Dmodel: online taxicab demand model from big sensor data in a roving sensor network, *available* <http://www.cis.temple.edu/~slin/Publications/ODmodel.pdf>.

[19] Shoham, Y. and Leyton-Brown, K., Multiagent Systems: Algorithmic, Game-Theoretic, and Logical Foundations, Cambridge University Press, (2009).

[20] Karl Bury, Statistical Distributions in Engineering, ISBN: 9780521635066, 1999

[21] Geoffrey Grimmett, David Stirzaker, Probability and Random processes: Third edition, Karl Bury, Statistical Distributions in Engineering, ISBN: 0 19 857223, 2001

[22] Han-wen Chang Yu-chin Tai Hsiao-wei Chen Jane Yung-jen Hsu, Context-aware taxi demand hotspots prediction. 2010, *International Journal of Business Intelligence and Data Mining*

[23] Han-wen Chang Yu-chin Tai Hsiao-wei Chen Jane Yung-jen Hsu iTaxi: Context-Aware Taxi Demand Hotspots Prediction Using Ontology and DataMining Approaches. 2008, *In Proceedings of the 13th Conference on Artificial Intelligence and Applications*

[24] Siyuan Liu, Yunhuai Liu, Lionel M. Ni, Jianping Fan, Shenzhen, Minglu Li, Towards Mobility-based Clustering, 2010, *Proceedings of the 16th ACM SIGKDD international conference on Knowledge discovery and data mining*

[25] Yan Huang, Jason W. Powell, Detecting Regions of Disequilibrium in Taxi Services Under Uncertainty, 2012 *SIGSPATIAL '12*

[26] Castro, P. S., Zhang, D., Li, S., Urban traffic modelling and prediction using large scale taxi GPS traces, 2012 *In Pervasive Computing (pp. 57-72). Springer Berlin Heidelberg*

[27] Phithakkitnukoon, S., Veloso, M., Bento, C., Biderman, A., Ratti, C., Taxi-aware map: Identifying and predicting vacant taxis in the city, 2010 *Ambient Intelligence (pp. 86-95). Springer Berlin Heidelberg*

[28] Bilton Nreick Disruptions: Taxi Supply and Demand, Priced by the Mile, 8.1. 2012, The New York Times.

[29] What is DRT?, <http://www.drtbus.co.uk>

[30] Oliver D. Hart, The Market Mechanism as an Incentive Scheme, The Bell Journal of Economics, 1983

[31] Hai Yang a, S.C. Wong b, K.I. Wong b, Demand–supply equilibrium of taxi services in a network under competition and regulation, 2011

[32] Matej Grajciar, Dr. Malcolm Egan, Martin Schaefer and Michal Jakob, Where Should Taxis Wait for Passengers in On-Demand Transport Markets?,

[33] Michal Jakob, Zbynek Moler, Antonin Komenda, Zhengyu Yin, Albert Xin Jiang, Matthew P. Johnson, Michal Pechoucek and Milind Tambe, AgentPolis: Towards a Platform for Fully Agent-based Modeling of Multi-Modal Transportation (Demonstration), 2012, 12th International Conference on Autonomous Agents and Multiagent Systems

[34] Uber, <http://www.uber.com>

[35] Lyft, <http://www.lyft.com>

[36] Liftago, <http://www.liftago.com>

[37] 18th IEEE International Conference on Intelligent Transportation Systems, September 15-18, 2015. , <http://www.itsc2015.org/>

[38] Flexible Mobility Services Testbed, <https://github.com/agents4its/mobilitytestbed>

[39] Dan Kedmey, This Is How Uber's 'Surge Pricing' Works, 15.12. 2014, Time magazine, <http://time.com/3633469/uber-surge-pricing/>

Chapter 8

CD content

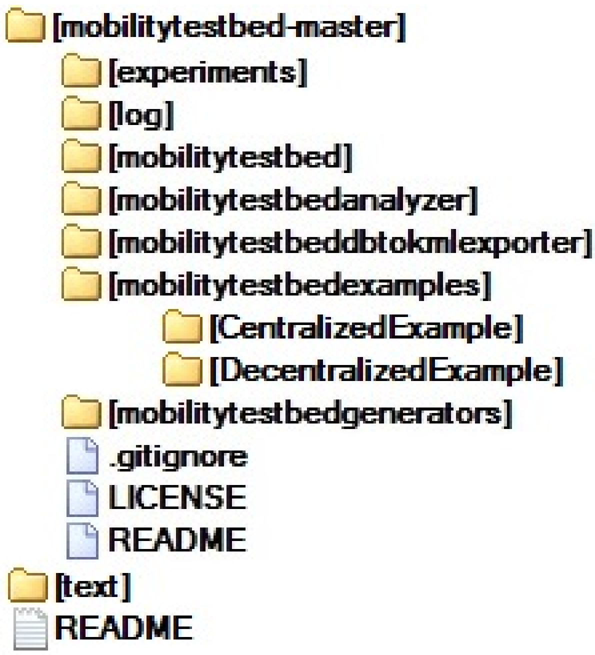

Figure 8.1: CD content

1. **mobilitytestbed-master**: contains the packages for the testbed simulation.
2. **experiments**: scenario folder.
3. **CentralizedExample**: contains the code with recommendation system.
4. **text**: contains the pdf of the thesis.