# Integrating Pedestrian into Ride-hailing Routing: The Potential Benefits

Yunhan Liu, C23025395,

Superviser: Dr. Padraig Corcoran,

MSc Computing

School of Computer Science and Informatics, Cardiff University

September 6, 2024

# Acknowledgement

The author would like to express his sincere appreciate to his supervisor Dr. Padraig Corcoran for his guidance and encouragement. During this project, Dr. Padraig Corcoran shared many experiences and skills, which made this project successfully completed.

# Contents

# Abstract

In traditional ride-hailing routing, the shortest driving distance from the pickup point to the drop-off point is usually prioritized. However, this may cause possible walking shortcuts to be ignored, and the hailed vehicles are forced to take detours.

Based on this fact, this study proposes a new pick-up and drop-off point optimization and routing algorithm based on the bidirectional A* algorithm, which utilizes the time passengers wait for ride-hailing vehicles to take a walk and takes the total travel time as a constraint to bypass the detour in the city. The algorithm is expected to reduce passengers' time in the car and fare without spending much cost and avoid online ride-hailing vehicles from entering congestion, one-way streets and unnecessary detours. For the online ride-hailing industry, it can improve vehicle operation efficiency, expand the potential users, reduce mileage and carbon emissions, enhance industry competitiveness and promote sustainable development of the industry.

To validate the effectiveness of the algorithm, this study uses taxi datasets in three cities: New York, San Francisco, and Chengdu for simulation experiments. The results show that in most cases, the algorithm provides a success rate of more than 90% within an acceptable running time. For the 2% to 7% of the total trips with detours, the average mileage can be reduced by 23% to 30% and the total travel time by about 2% to 10%. The mileage of the car-hailing industry can be reduced by 0.2% to 1.2%. In addition, the study also speculates that different urban structures and urban planning will have a significant impact on the detour rate, mileage reduction and other figures related to vehicle detours.

# 1 Introduction

## 1.1 Background and Motivation

With the development of cities and transportation systems, ride-hailing services have been an integral mode on daily transportation [23] due to its privacy and convenience. Passengers can customize their pickup and drop-off locations through online ride-hailing services for a personalized door-to-door travel experience, which provides great convenience for passengers [19]. Most of these ride-hailing services are based on online applications. By using real-time network communications, ride-hailing services can effectively and timely match empty vehicles with nearby user demands [33], making them have a much better performance than traditional street-hailing taxis in terms of reliability, spatial coverage, and operating costs [46].

Ride-hailing services have achieved an impressive market penetration in this decade. By 2023, Didi had 587 million annual active users and 23 million annual active drivers in 16 countries since 2012. In China, Didi perform averagely more than 28 million daily transactions including ride-hailing, ride-pooling, etc [16]. And Uber has 150 million monthly active platform consumers and provides 2.601 billion trips per year [55].

However, one of the potential problems is the increased milage due to unavoidable long detours on driving routes. These detours are caused by bypassing some congested block or common city facilities and layouts such as parks, pedestrian streets, railway tracks, and rivers. These points in the trip may be very close in a straight line, but the driving distance is very long. In navigation strategies, usually, pedestrian pathways are not considered. One approach could reduce and avoid detour could be incorporating short distance walking to determine a new flexible and ideal location for pick-up or drop-off. For passengers, a short-distance walking brings less time in the car and less fares. For car-hailing drivers and operators, it can avoid detours, save fuel and mileage, and improve operational efficiency.

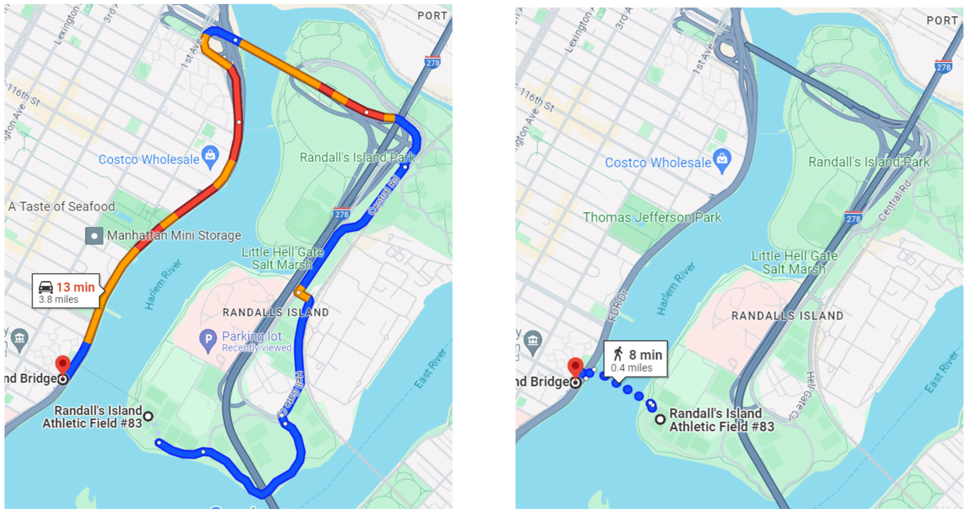
Furthermore, the ride-hailing industry has an urgent demand to reduce their side effect on the environment and traffic. It is reported that the excessive use of ride-hailing services may lead to increased vehicle mileage [28], exacerbate traffic congestion and cause excessive energy consumption and air pollution [54] [56]. Improving the operational efficiency of the ride-hailing services is crucial to reducing these negative externalities [19].

This project aims to provide a potential way to reduce overall travel costs by providing passengers with suggestions and options for alternative pick-up or drop-off points. Specifically, based on the analysis of the low efficiency of car-hailing trips, a new pick-up and drop-off point selecting strategy based on A-star algorithm will be proposed, measuring by time including walking, driving, taxi waiting and extra walking time. Additionally, three datasets in New York, San Francisco, and Chengdu will be used to evaluate the benefits of the algorithm.

## 1.2 Problem Statement

In traditional ride-hailing service apps or map apps, the shortest driving distance from the pick-up point to the drop-off point is usually prioritized. However, this may let possible walking shortcuts being ignored and hailed car forced to take a long detour. Here, the detour here can be broadly defined. It can refer to common city facilities and layouts like parks, pedestrian streets, railway tracks, and rivers, or it can refer to a generalized defined detour like areas that are inconvenient to enter, one-way neighbourhoods and congested areas, or even to let passengers go across the street to take up the vehicle.

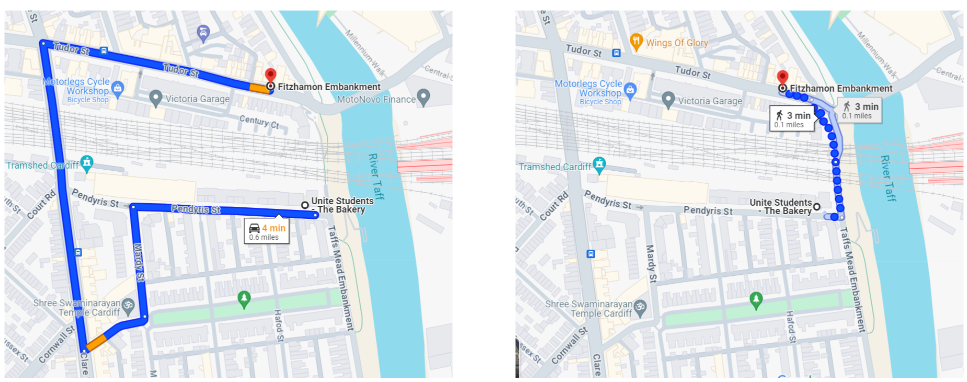
For passengers, the detour increasing the total travel time and money costs. An ultimate example would be like "crossing a pedestrian bridge over a river" to hail a taxi rather than hail it here and have the vehicle detour around the river. In existing solutions, traditional map applications offer "mixed mode" options, but usually, options are very limited and usually do not contain an option targeting on optimize ride-hailing services.




(a) Driving route given by Google map (b) Pedestrian bridge over the river

Figure 1: An example on potential detour for passenger

And for drivers, the passenger's pickup or drop-off point may be in an inconvenient location, which may include a one-way street or a congested block. It takes a lot more time and fuel for vehicles to enter or leave these areas than average. Passengers may only need to walk a short distance to avoid these unpopular areas.




(a) Driving route (b) Walking path

Figure 2: An example on potential detour for driver

For the sustainable development of online ride-hailing services, it is reported that if walking is integrated into ride-hailing services, the rejection rate can be reduced by more than 80% and the mileage can be reduced by more than 10%. [33]. Without walking, ride-hailing trips may even be less efficient than traditional street-hailing taxis [39].

In order to mitigate the detour and solve the problem, specifically, the following problems need to be addressed:

- How can urban road networks, containing both pedestrian pathways and vehicle roads, be effectively mapped and modeled to optimize taxi routes?
- How to check the existence and availability of walking shortcuts in real-time navigation routes?
- How to measure the effectiveness of optimizing taxi and pedestrian integrated routing systems to ensure optimal travel time and cost without compromising convenience?

The result of this project is expected to be valuable to both ride-hailing drivers and passengers. By utilizing the waiting time for a ride, or a few extra minutes to introduce walking, and selecting new pick-up and drop-off points, passengers can expect to reduce fares without losing much time, or even saving more time. Online car-hailing drivers can avoid driving into congested areas or entering unnecessary detours. For the sustainable development of the ride-hailing industry, it can improve operational efficiency, expand the potential user base, reduce mileage and carbon emissions, and make online car-hailing more competitive in the face of other modes of transportation.

## 1.3 Aims and Objectives

The aim for this project is to explore the potential benefits of moving pick-up and drop-off points within a certain walking distance to improve the efficiency of online ride-hailing routes and devise a navigation approach on identifying optimal pick-up and drop-off points by reducing overall cost. Specifically, the project has the following objectives:

- Develop on mapping system for processing routing information, with pedestrian pathways, traditional road networks, and important environment features integrated together.

- Design a mechanism, which can dynamically check the existence and give pick-up and drop-off point of walking shortcuts in routing and providing walking-driving mixed route options.
- Use 3 real-world trip datasets to test the effectiveness and potential benefits of the new approach.

Also, long-term goals include collecting data and parameters from various cities, identifying possible obstacles in advance to optimize operational efficiency of the approach, working with map service providers to obtain more accurate road network and navigation data, and cooperating with online ride-hailing service providers to implement the project into actual routes and navigation.

## 1.4 Outline of the Report

The remainder of this project is organized as follows.

- Section 2 will review some prior studies to understand the development of ride-hailing industry and existing solutions to the pick-up and drop-off problem.
- Section 3 will introduce some basic concepts, definitions and theoretical foundations for the research.
- Section 4 will explain the inefficiency in ride-hailing trips and provide a new approach for integrating walking into ride-hailing trips.
- Section 5 will use three real-world trip datasets from difference cities to test the effectiveness and potential benefits of the new algorithm.
- Section 6 will conclude this project and provide some reflections and future works.

# 2 Existing Literatures and Solutions

In this section, prior studies will be introduced and reviewed to demonstrate the features and development of ride-hailing service and illustrate why integrating walking into ride-hailing has a promising future. Additionally, some research gaps will be identified.

## 2.1 Development of Ride-hailing Services

As an innovative practice, ride-hailing is defined as a kind of service that can connect drivers and customers in real-time and enable customers to take rides through online platforms to reach their destinations[12]. Benefited from the reliability and convenience provided by advances in network and communication technologies, they got remarkable and rapid growth in the transportation market [54]. For example, by 2018 [13], Didi provides their service to approximately 450 million users together with 21 million drivers since 2012, and performs over 30 million trips in more than 400 cities in China every day. In San Francisco, approximately 3% of the city traffic shares are performed by ride-hailing services including Uber and Lyft, which is about 12 times more than traditional taxi services, according to the data provided by SFCTA [38].

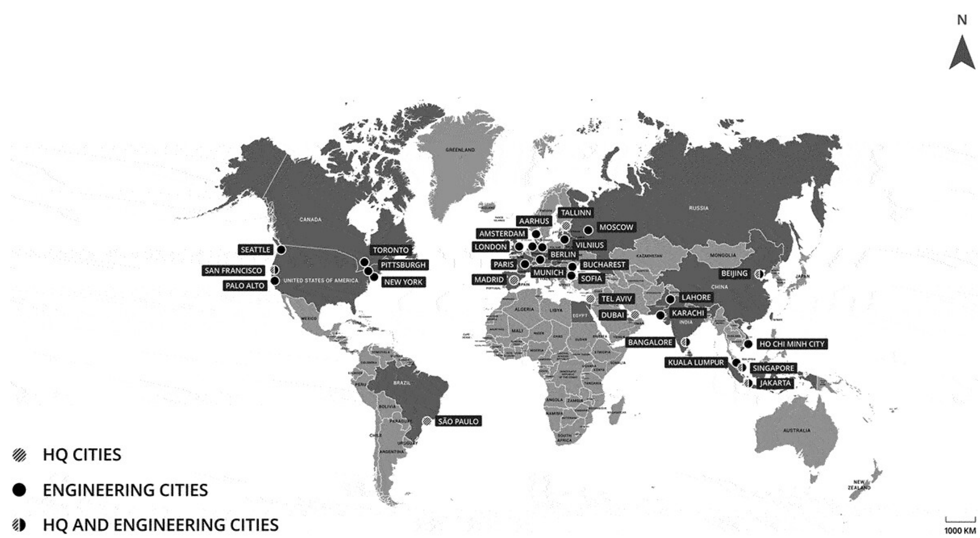



Figure 3: Major cities using ride-hailing services [7]

### 2.1.1 Who and Why Use Ride-hailing

To completely understanding the role ride-hailing service plays, one of the necessary questions is user characteristics. Generally, ride-hailing services is more widely used among young, well-educated and wealthy people. According to the studies in Southeast Asia

[11], most of passengers are under 25 years old. This figure in United States is the age between 25-34 [24]. In United States [24], 70% of the users have a bachelor's degree. Similar circumstances have also appeared in Europe [23], with more than half of users are students. However, in China [51], it is less related to education, with most of users not having received higher education. Also, people who are more familiar with modern technologies are more likely to become the users of ride-hailing services mentioned by [1], [23] and [24].

Regarding the use of online ride-hailing services, usually, large cities and high-density areas are generally considered to be areas where online ride-hailing services are more frequently used [2] [23] [24] [51], with relatively short trips of less than 30 minutes [11] [51]. Additionally, in terms of purposes, studies show that in China [51], commuting and entertainment are the main reasons for online ride-hailing trips, accounting for 40% and 33% of the total number of trips respectively. While in some Southeast Asia countries [11], the majority of ride-hailing trips are for either leisure or social, similar to Rayle et al.'s finding [45] in San Francisco, United States and Tirachini & Rio's finding [53] in Santiago, Chile.

Interestingly, there is an argument on the relationship between car ownership and the level of online ride-hailing services. In United States [1], compared to non-frequent ride-hailing service users, a broad spectrum of frequent users live in car-free households. However, a different situation may be the case in other part of the world, here, more than two-thirds of the ride-hailing service users in China are private car owners [51], and this figure ranged from 22% to 65% in some of Southeast Asia countries [11] and from 45% to 48% in Chile [53]. For car owners, in addition to the common purposes mentioned above, additional reasons for using online ride-hailing services include the inconvenience of parking in the city centre [23] [24] [51], not having to drive after drinking [54], and road restrictions on private cars [51].

In short, the user group of online car-hailing is mainly wealthy young people living in big cities, which means it has great development potential. And the reduction in fares might attract more people to use the service. Many of the users own private cars, which means that if the industry develops, it is possible to reduce the use of private cars and reduce the impact on traffic and the environment.

### 2.1.2 Substitution and Complement to Other Modes

Ride-hailing services are commonly considered as an alternative and / or supplement to other modes of transportation [54]. In all studies, taxis and public transport are the most commonly discussed modes of transport that replace / complement ride-hailing services [54].

#### Traditional Taxi

Traditional taxis are seen as facing fierce competition from online ride-hailing services [51]. Many researchers have observed that the impact of online ride-hailing on traditional taxis is mainly a substitution effect, that is, a decrease in traditional taxi passengers [14] and a decrease in taxi drivers' income [5]. This may be because online ride-hailing services and traditional street-hailing are considered to have similar market demand and both services cover similar areas and trip length [45]. Compared with traditional taxis, online ride-hailing services are more reliable. Specifically, according to the Brown and LaValle [8], 99.7% of ride-hail trips were guaranteed of reaching their destination, while 19.5% of traditional taxi passengers were not picked up. Also, Riding-hailing services usually provided lower waiting time, with 5.7 minutes on average. By contrast, averagely, traditional taxi passengers spend four times as long waiting, at 24.3 minutes [8]. Similar phenomenon appears in Chalermpong 's report [11].

In addition to reliability, price is another factor that makes online ride-hailing more popular than traditional taxis. In a field study in Los Angeles, traditional taxi trips were 40% more expensive than ride-hailing services averagely and sometimes it can be "double the price" compared with online ride-hailing [8]. In China, the general situation is similar, with the slight difference that as travel time increases, the price advantage of online ride-hailing services will become less significant [51]. Also, transparent fares and fewer detours make passengers more satisfied with online ride-hailing services [8].

### Public Transportation

In terms of public transportation, ride-hailing services can be treated as both alternative and supplement [54]. On the one hand, most of public transportation have fixed routes and schedules. Ride-hailing services can complement public transportation by filling in theses gaps [26]. This model of combining public transportation and ride-hailing for longer trips is also known as the first-mile or last-mile mode [54]. According to Cats et al.'s research [10] in 6 cities, 20% to 40% of ride-hailing trips have no available public transit alternative. There are also other studies in United States [24] and Chile [54] show that the use of public transportation is positively correlated with the use of online ride-hailing services. Besides, urban planning itself will also have an impact on the supplementary effect of online ride-hailing services. Cats [10] pointed out that the improvements in accessibility (measured by travel time) brought about by ride-hailing services are particularly significant in areas not yet covered by urban rail transit, with a noticeable example of US cities generally having a higher added value of accessibility than European cities.

On the other hand, compared to fixed-line public transportation, ride-hailing provides more convenience, comfort, and flexibility to passengers [62], which makes it more competitive compared to public transportation, thus replacing it. Research by Lee et al. [32] revealed that Uber has decreased the need for public transportation in most developed nations, particularly in densely populated cities. Also, 9% to 15% of Uber trips have potential public transportation alternative with a shorter ride [10]. Qiao and Yeh used Chengdu as case study and conclude that ride-hailing services are attracting people, especially low-income families, away from public transportation. In addition, if ride-hailing services are not available, 38% of ride-hailing orders may use public transportation, which ranks first among all options, higher than traditional taxis (36%) and private cars (26%) [44]. What's worse is that in long-term, public transportation ridership declining could lead to public transportation service cuts and reductions, increasing wait and travel times for their users [54].

As discussions on substitution and complementary relationship between online ride-hailing services and other mode of transportation, it can be found that taxis, as the main substitute for online ride-hailing, have many similar features to online ride-hailing, which means that the literature on taxis can provide experience for investigating online ride-hailing issues. The substitution and complementary relationship between public transportation and online ride-hailing can also provide theoretical support for integrating walking into online ride-hailing routing.

### 2.1.3 Prospect and Sustainable Development

Many studies have focused on the effect of ride-hailing services on environment and traffic, one of the problems is vehicle kilometre travelled (VKT) or vehicle mile travelled (VMT). Different articles hold different views on whether online ride-hailing will increase VKT. Although some of the studies implied that ride-hailing services will not significantly increase VKT [1] [39], there are still a lot of studies showing that ride-hailing services will increase VKT and have side effects on the environment. Henao and Marshall's study [28] examined hundreds of ride-hailing trips in the United States and innovatively used empty-load rate and vehicle occupancy rate as evaluation indicators. They claimed that although ride-hailing is more efficiency than traditional taxis and private car in terms of transportation passengers per VKT, ride-hailing services provided an extra portion (83.5%) of VKT to the system when considering empty trips, induced trips, and more sustainable transportation alternatives [28]. Similarly, Wenzel [56] analysed the data from 1.5 million ride-hailing trips in Austin and found that for trips within 60 minutes, drivers are traveling averagely 55% more distance between drop-off and next pick-up, which led to 41% to 90% increase on ride-hailing trip energy consumption compared with previous mode. Besides, by analyse the orders of Didi in 2015, Wu et al. [59] also revealed that ride-hailing services in China have limited VKT substitution to private cars (43%) and ride-hailing services caused around 0.25 million tons of extra gasoline consumption, which contributed to 0.79 million tons of extra $CO_2$ emissions every year.

Due to the high VKT of ride-hailing services, to promote sustainable development of ride-hailing services, another topic is electrification. The use of electric vehicles is considered as a practical option to reduce energy consumption, traffic pollution and greenhouse gas emissions [20]. Slowik and his team [49] investigated the adoption of electric vehicles by the top five global ride-hailing companies in 2019, found that electric vehicles accounted for only a small percentage of ride-hailing vehicles. Due to the huge potential for the electrification of online ride-hailing vehicles, many authorities and companies have adopted various strategies. In areas such as California, London, and Shenzhen, authorities are drafting regulations to steer ride-hailing companies towards electric vehicles [50]. Industry giants like Didi and Uber have also set different goals on future fleet electrification [49].

On the other hand, ride-sharing is expected to reduce VKT and promote the sustainable development of ride-hailing services. It has been shown that ride pooling can improve urban mobility, provide greater convenience, flexibility and comfort than fixed-route public transport, while offering significantly lower costs than individual rides [62]. Studies also show that for the giant cities like New York, Beijing and Shanghai, more than half of the trips can be shared [9] [47] and reduce VKT by 33% to 40% [9] [47], 15% to 23% of fuel consumption [60] and averagely 23% of polluting emissions [60]. Ride sharing is also seen as a promising option to solve the last kilometre problem [62]. However, research on ride sharing tends to more focus on large cities who have more demand, and thus, a greater likelihood of similar origin-destination pairs [63]. Furthermore, the penetration rate of ride sharing remains low [28], which may be due to the low probability of successful matching [28] or excessively transfer cost to passengers, forcing them to increase overall travel time [34].

In the discussion about the sustainable development of online ride-hailing, there is a lack of mature solutions, hence the discussion on the optimization of pick-up and drop-off locations is promising. The choice of pick-up and drop-off locations might affect travel efficiency, congestion and carbon emissions [19], so it is worth to explore the potential impact of pick-up and drop-off locations on online ride-hailing.

## 2.2 Integrating Walking into Mobility System

Asking passengers to take walks might be easier than expected and able to generate relevant advantages to the system [22]. Although ride-hailing services usually have much shorter waiting times than traditional taxis [8], it still has the potential to be exploited. Fielbaum et al. [22] claimed that concerns toward inconvenient walking can be rewarded by reducing waiting and in-vehicle time. Also, some passengers in metropolitan would take their time walking rather than waiting [31]. Besides, many studies [11] [45] suggested that most ride-hailing trips take less than 10 minutes waiting time, and the average waiting time is 4 to 5 minutes [61]. These facts may indicate that the 4 to 5 minutes waiting time [61] can be changed into walking without gaining too much inconvenience to passengers.

Walking plays an important role in many other transportation modes. In terms of traditional taxis, vacant taxis usually tend to cruise on the streets looking for their next passenger or drive up to passengers waiting in line at taxi stands [58]. Taxis cruising the streets usually search for passengers on the main or high-level roads, which means that passengers also need to walk there to hail a taxi [19]. Also, cruising taxis usually tend to look for passengers in densely populated areas, commercial areas, and even some congested areas [58], which means that for sparsely populated areas, the demand for walking may be greater than expected. On the other hand, taxis that choose to go to the taxi stand might reduce fuel consumption [58]. But this means that passengers will have to walk to the taxi stand and wait in line to get a taxi. In contrast, ride-hailing services, which share many similarities with traditional taxis, have lost lots of the flexibility offered by walking due to their overemphasis on the convenience of door-to-door service [19]. Without walking, ride-hailing trips may even be less efficient than traditional street-hailing taxis [39].

Walking is also needed for most public transportation [22] [30] [37]. The first / last mile problem in public transportation has been discussed for a long time. For traditional buses and subways with fixed stops, their appeal to residents depends largely on whether people can walk to the nearest stop [30]. Their stop design has already implied the walking part of arriving at the station and transferring. For some new types of public transport, like on-demand minibus services with flexible routes and schedules, or customized buses, although their station design can reduce walking, in order to reduce carbon emissions and congestion [35], as well as for profit purposes, passengers are also required to take a short distance walking to a fixed meeting point [19].

Until recent years, research exploring the integration of flexible pick-up and drop-off points with ride-hailing services has been limited, and most studies have focused on ride-sharing services. Studies by Fielbaum and his colleagues in 2021 [22] and 2022 [21] proposed and improved the vehicle routing and pick-up and drop-off points problem in on-demand ridesharing, and used heuristic algorithms to solve related optimization problems, filling the gap in this field. Also, the study by Aliari and Haghani [3] used a 2015 New York City yellow taxi dataset to form clusters of passengers with high carpooling potential and incorporated these alternatives into a novel mixed integer linear programming (MILP) formulation to find the optimal schedule, which ultimately reduced waiting time by about 50% and reduced total trip time and vehicle used.

In terms of integrating walking into standard car-hailing systems rather than car-pooling, Megantara et al. [36] conducted a numerical simulation based on the data of Manhattan New York City taxi trip in 2013, showed that within maximum 3 minutes delay for every travel, approximately 12% of fuel, which can support 800 kilometers trip, was saved during 1 hour with all users willing to take 300 meters walk. Sarma et al. [48] compared ride-hailing and ride-pooling for both traditional and including walking groups. By analyzing historical data, it has been proved that there is a 5 to 10% reduction on the average vehicle kilometer traveled per served request. By examining inefficiency in ride-hailing trips, Ding [19] designed a two-stage heuristic approach by minimizing detours and avoiding congestion. Their team conducted analysis on Didi data from Chengdu downtown and proved that about 8% to 23% of trips can be optimized, which finally leads to up to 15% reduction on travel distance, energy consumption and carbon emissions.

## 2.3 Research Gap

In summary, optimizing the pick-up and drop-off locations for ride-hailing services is expected to improve travel efficiency, reduce congestion and carbon emissions, and promote the sustainable development of the industry. The integration of walking and other modes of transportation has been proven to be beneficial, and integrating walking with ride-hailing services also has great potential.

However, with the development and changes in the economy and society, people's living habits have changed a lot, especially after the pandemic, which causes the ride-hailing industry has grown rapidly, which means that many of the finding concluded from older literature are no longer accurate. Furthermore, studies mainly focus on a limited number of large cities with large and concentrated populations like San Francisco, New York in United States and Beijing, Shanghai, Chengdu in China. Many other large cities, as well as numerous smaller ones, have been ignored. It is a fact that studies in a limited number of cities still have conflicting outcomes. These facts may imply that different times, cities, and policies may lead to completely different results. All of these can be proved by the conflicting findings and conclusions in previous studies conducted in different years. Therefore, it is difficult to say that the literature above is highly representative and can be applied to any city.

Apart from that, data on ride-hailing services is mostly owned by a few large, competing industry giants and involves trade secrets and commercial interests. Therefore, most trip data is not public. In this circumstance, research that relies on merely limited datasets and user recalling on their trips is likely to have different degrees of bias. And there are very few literatures that can obtain accurate original travel data for research. This also explains why the research is concentrated in a few specific large cities.

Although many studies have examined ride-hailing operating strategies, most of them don't pay enough attention to the potential impact of pick-up and drop-off locations, which is typically overshadowed by focusing on the convenience of door-to-door services [19]. Limited existing literature that incorporates pick-up and drop-off points are mostly focused on the ridesharing scenario [19]. To address these research gaps, this project will devise a navigation approach on optimizing ride-hailing vehicle pick-up and drop-off point by incorporating walking to reduce overall cost.

# 3 Concepts, Definitions and Theoretical Foundations

Before introducing the algorithm, some concepts, definitions and theoretical foundations will be firstly introduced in this section.

## 3.1 Graph Theory and Street Network

In this project, the street network will be modelling as graph, where streets will be abstracted as edges and crosses will be abstracted as nodes, to find potential optimal pick-up and drop-off points.

### Graph

Obviously, the actual street network contains length and direction. According to the Diestel [17], a directed graph with weight can be denoted as graph $G(V, E, W)$ where $V$, $E$, $W$ denotes vertices, edges and weights sets correspondingly. For every edge $e$ in set

$E$, there is an initial vertex and a terminal vertex, and the edge $e$ is said to be directed from initial vertex to terminal vertex, with the weight of $w$. In street network modelling, vertices are referred as nodes and weights are referred as length or distance of the street. All the edges are expected to have a positive weight.

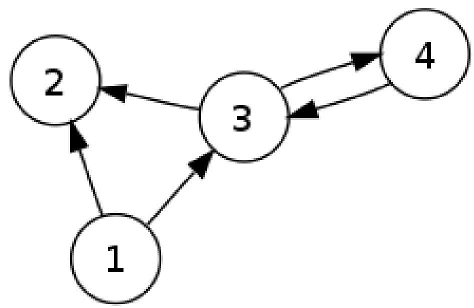



Figure 4: A simple directed graph [57]

**Street Networks**

The street networks are retrieved from OpenStreetMap (OSM). OpenStreetMap is open data, licensed under the Open Data Commons Open Database License by the OpenStreetMap Foundation [40]. It contains many information that can be utilised in this project. The implement of this project will retrieve street network by type such as walking, driving and utilise length in street properties as weight of the edges.

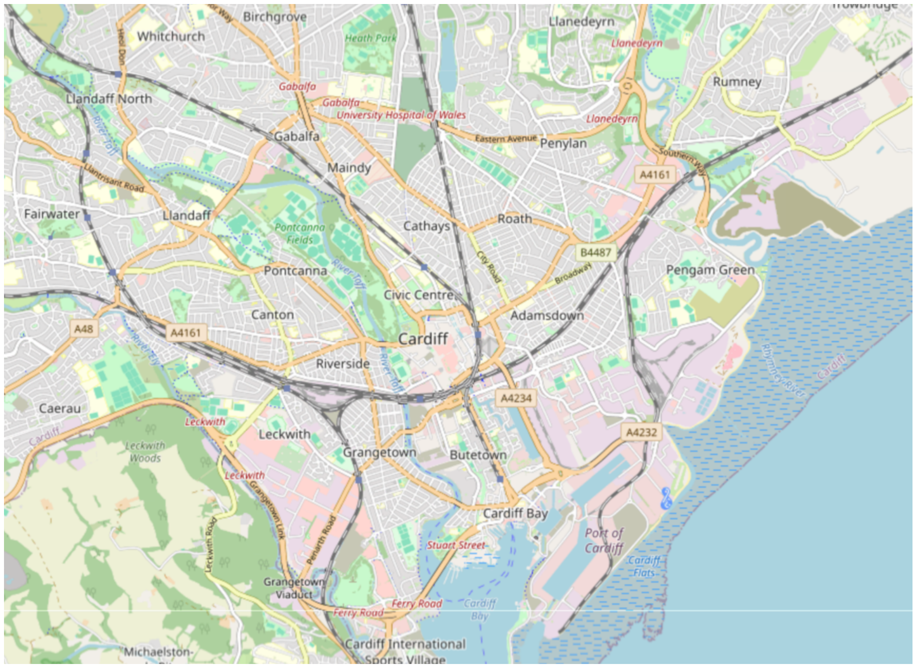



Figure 5: An example of OpenStreetMap

**Coordinates, Projections and Conversions**

Since the earth is an ellipsoid rather than a sphere, projection is necessary for mapping and modelling the surface of the earth. The coordinate system OpenStreetMap uses is WGS-84 [41], which means that for accuracy, coordinates must be projected before process and analysis, then fed into the program of street network for further computation.

## 3.2 Graph Searching Algorithms

Graph search algorithms can find the shortest path on a map represented by a graph. Their development includes the following three important algorithms: Breadth First Search (BFS), Dijkstra's algorithm, and A* algorithm.

### 3.2.1 Breadth First Search (BFS)

The key idea of BFS is not complicated. What BFS does is to set up a queue, put the starting point into the queue, and each time an element is taken out of the queue, determine whether it is the end point. If not, check all the neighbours of this element and add any passable neighbours that are not the end point to the queue. Repeat this process until the queue is empty.

The Algorithm follows these steps:

---

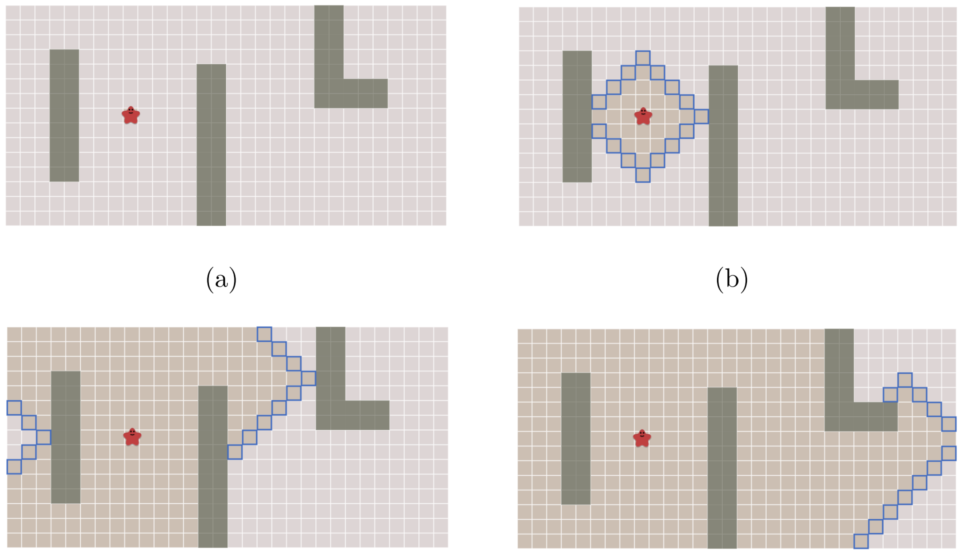
**Algorithm 1:** Breadth-First Search (BFS) [4]

---

```
queue ← Queue()
queue.put(start)
came_from ← {}
came_from[start] ← None
while queue.empty() = False do
    current ← queue.get()
    if current = end then
        break
    for next ∈ graph.neighbors(current) do
        if next ∉ came_from then
            queue.put(next)
            came_from[next] ← current
```

---

The following figure shows the searching process of BFS, where the pink star represents the start point, the blue area is the search boundary, and the dark grey area represents the inaccessible area.



(a) (b) (c) (d)

Figure 6: Process of Breadth First Search [4]

### 3.2.2 Dijkstra's Algorithm

Usually, in real street networks, each street has a cost (distance), which should be considered in street network modelling. BFS is based on the number of steps to expand the search boundary. In this case, the search should be based on distance instead of number of steps. And this is the basic idea of Dijkstra's algorithm.

The following figure shows the difference between step-based search and cost-based search, where the dark grey area is set to be impassable, and the cost of the green area is set to five times that of the normal white area.

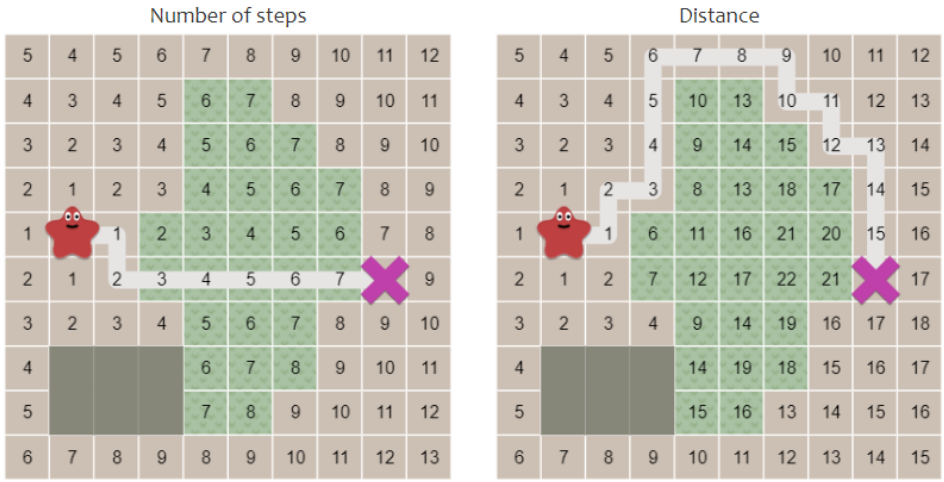


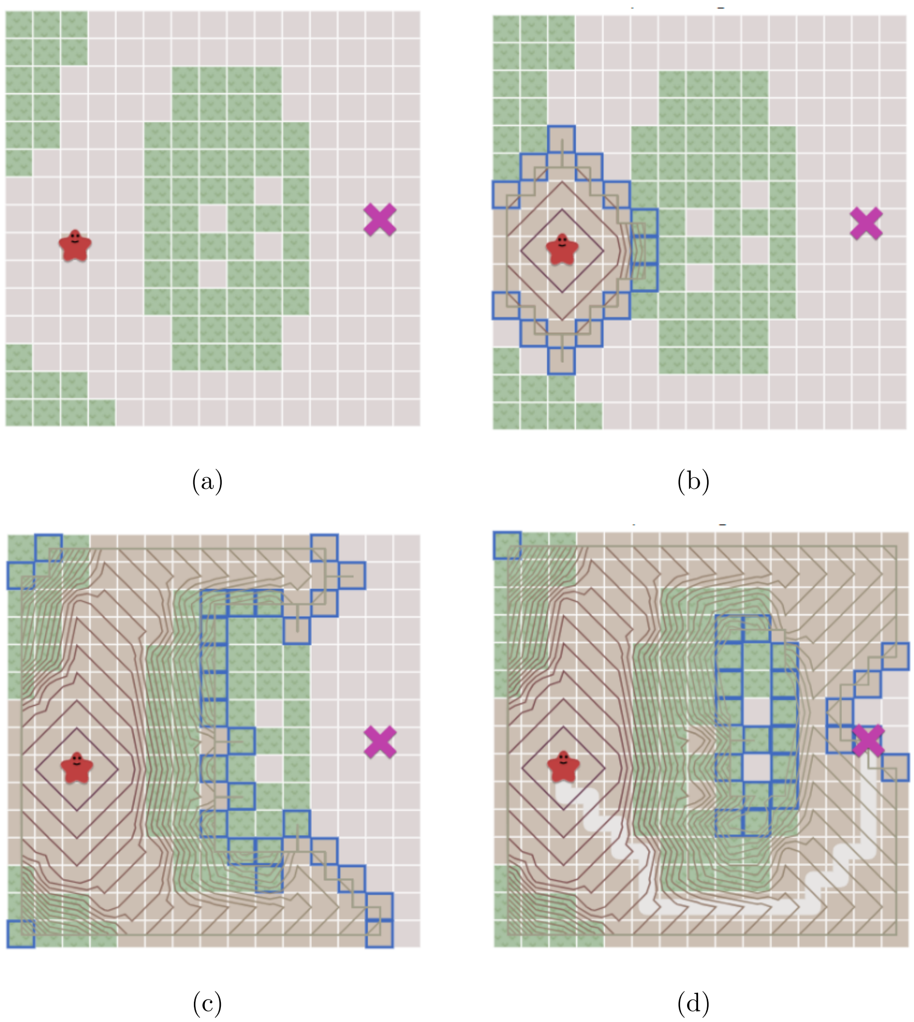

Figure 7: BFS vs Dijkstra's Algorithm [4]

Dijkstra's algorithm was proposed by E. W. Dijkstra in 1959 [18]. The key idea of Dijkstra's is similar to BFS while it takes cost (distance) into account. It maintains a priority queue of nodes, sorted by distance, and a "distance field" used for storing total movement cost from the start point. The priority queue initially has only start point, with the cost (distance) of 0. The existence of the priority queue ensures that each point taken out has least cost (closest to the starting point), which ensures that the path is the shortest. And each round of loop will take out the first element from the top of priority queue and determine if it is the end point. If not, go through all the neighbours of current point and calculate the distance travelled so far. If this distance so far is less than the distance storing in the "distance field", the distance in "distance field" will be refreshed and this neighbour will be added into priority queue. This process will be repeated until path is found or the priority queue is empty.

**Algorithm 2:** Dijkstra's Algorithm [4]

```
frontier ← PriorityQueue()
frontier.put(start, 0)
came_from ← {}
came_from[start] ← None
cost_so_far ← {}
cost_so_far[start] ← 0
while frontier.empty() = False do
    current ← frontier.get()
    if current = end then
        break
    for next ∈ graph.neighbors(current) do
        new_cost ← cost_so_far[current] + graph.cost(current, next)
        if next ∉ cost_so_far or new_cost < cost_so_far[next] then
            cost_so_far[next] ← new_cost
            priority ← new_cost
            frontier.put(next, priority)
            came_from[next] ← current
```

Similarly, the following figure shows the searching process of Dijkstra's algorithm, where the pink star represents start point and purple crossing represents end point. And the contour lines intuitively show the cost required to reach a specific grid.



(a) (b) (c) (d)

Figure 8: Process of Dijkstra's Algorithm [4]

### 3.2.3 Heuristic Search and A* Algorithm

Dijkstra's is already standard solution for many navigation packages. However, it is still not efficient enough. BFS and Dijkstra's algorithm explore in all directions, which is a reasonable choice when information near the start point is needed. However, a common situation is to find the shortest path from the start point to the end point. In this case, it is hoped that boundary can be extended from the start point to the end point instead of searching in all directions evenly with the start point as the centre.

A common algorithm that uses the distance to the end point as a measure is the greedy algorithm. It uses the Euclid distance or Manhattan distance from the current point to the end point as its heuristic function and sort the priority queue accordingly. With the same number of steps, the greedy algorithm is much more efficient than the Dijkstra's algorithm.

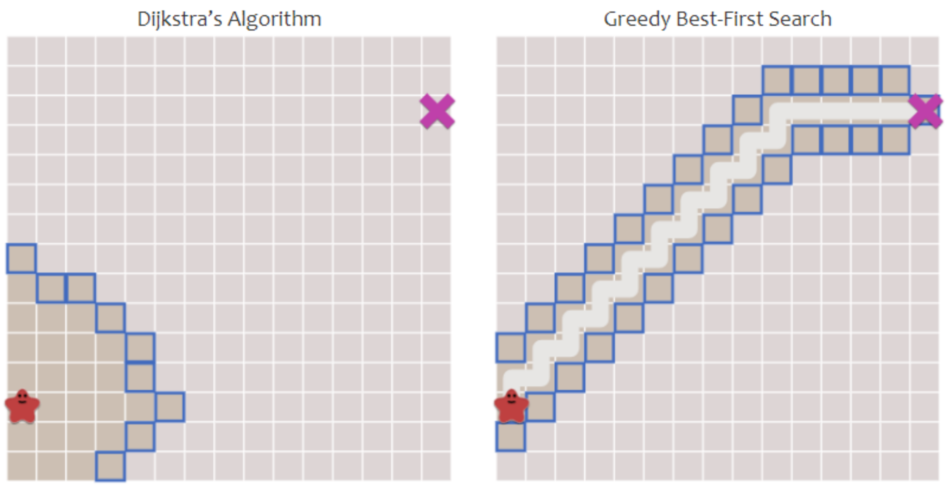



Figure 9: Dijkstra's Algorithm vs Greedy [4]

But when the map is more complex and contains more obstacles, the speed advantage of the greedy algorithm and Dijkstra's is shrinking, and greedy may provide a path that is not very good.

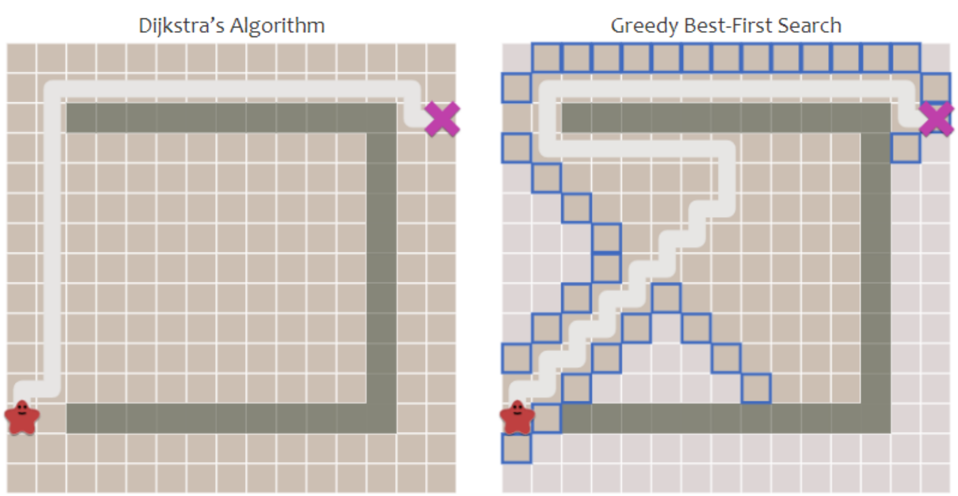



Figure 10: In more complex situations [4]

Dijkstra's algorithm keeps the path shortest but might waste steps on exploring unpromising directions. Greedy algorithm searches towards end point while might not output shortest path. A promising solution might be using both the distance travelled so far and estimated distance as heuristic function. And this is how A* algorithm works.

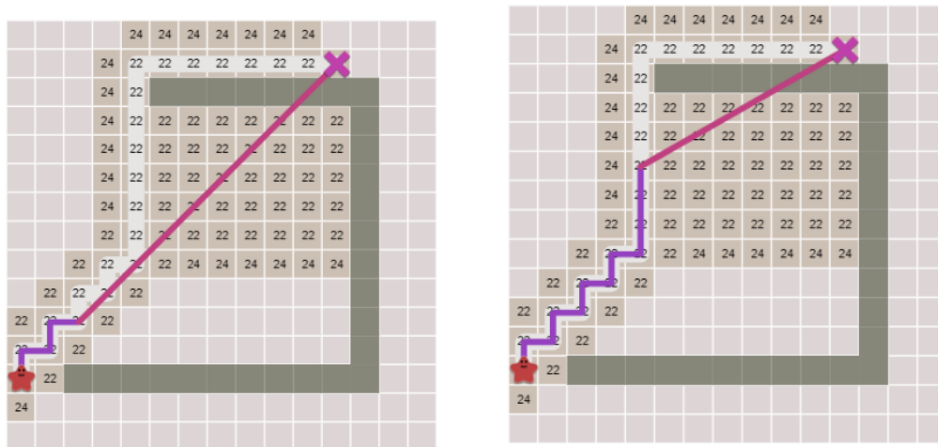

Figure 11: Difference between three algorithms [4]

Peter Hart, Nils Nilsson and Bertram Raphael first published the A* algorithm in 1968 [42]. It achieves better performance by using heuristics considering both distances travelled so far and estimated distance to guide its search, denoted by $f(N) = g(N) + h(N)$. It has very similar structure to Dijkstra's algorithm. It also maintains a priority queue of points, sorted by distance, and a "distance field" used for storing distances travelled so far. The priority queue initially has only start point, with the distance of 0. The existence of the priority queue ensures the path is the shortest. And each round of loop will take out the first element from the top of priority queue and determine if it is the end point. If not, go through all the neighbours of current point and calculate the distance travelled so far. If this distance so far is less than the distance storing in the "distance field", the distance in "distance field" will be refreshed and this neighbour, together with the sum of distances travelled so far and estimated distance, will be added into priority queue. This process will be repeated until path is found or the priority queue is empty.

**Algorithm 3:** A* Algorithm [4]

```
frontier ← PriorityQueue()
frontier.put(start, 0)
came_from ← {}
came_from[start] ← None
cost_so_far ← {}
cost_so_far[start] ← 0
while frontier.empty() = False do
    current ← frontier.get()
    if current = end then
        break
    for next ∈ graph.neighbors(current) do
        new_cost ← cost_so_far[current] + graph.cost(current, next)
        if next ∉ cost_so_far or new_cost < cost_so_far[next] then
            cost_so_far[next] ← new_cost
            priority ← new_cost + heuristic(end, next)
            frontier.put(next, priority)
            came_from[next] ← current
```

Similarly, the following figure shows the searching process of A* algorithm, where the purple line denotes distance travelled so far and pink line denotes estimated distance.



(a) (b)

Figure 12: A* algorithm (travelled so far in purple and estimated distance in pink) [4]

### 3.2.4 Bidirectional Searching

Till now, all graph search algorithms mentioned in this article are searching in one direction, usually from the start point toward the end point. And in terms of graph searching, another option is bidirectional searching, which contains two cursors and runs two simultaneous searches including forward searching from start point toward end point and backward searching from end point toward start point. This technique can reduce search time in many cases, especially in large graph, but it may also increase physical memory usage. And the new approach in this paper will be based on the bidirectional A* algorithm.

# 4 Integrating Walking into Ride-hailing Services

In this section, the inefficiency in ride-hailing trips will be revealed and explained, thus providing a theoretical basis for integrating walking into ride-hailing trips. Also, the assumptions, modelling and design of the new algorithm will be explained.

## 4.1 Inefficiency in Ride-hailing Trips

In order to integrate walking into ride-hailing services, firstly, the reasons for the inefficiency of online ride-hailing trips will be analysed based on the characteristics of online ride-hailing trips. So, actual trip datasets will be used for analysis.

### 4.1.1 Data Description

There are few public data sets on online ride-hailing trips. As online ride-hailing services and traditional street-hailing are considered to have similar market demand and both services cover similar areas and trip length [Rayle2016], for examining the effectiveness of the model, this study mainly utilized three opensource taxi datasets from New York, San Francisco and Chengdu to simulate the car-hailing trip dataset for analysis..

New York taxi trip dataset is based on prediction competition [29] hold by Kaggle. The competition dataset was originally published by New York taxi and Limousine Commission (TLC) and being sampled and cleaned for the purpose of competition. It contains 1458644 trip records with data fields including *id*, *vendor id*, *pickup datetime*, *dropoff datetime*, *passenger count*, *pickup longitude*, *pickup latitude*, *dropoff longitude*, *dropoff latitude*, *store and fwd flag* and *trip duration*.

San Francisco taxi dataset is provided by Piorkowski's team and available at IEEE open access dataset website [43]. It contains mobility trace represented by GPS coordinates of approximately 500 taxis from 17 May, 2008 to 10 June, 2008 in San Francisco. Each line of the dataset contains latitude, longitude, occupancy and time (in UNIX epoch format).

Chengdu taxi trip trajectory dataset is based on DataCastle data competition [15] hold in 2016. It contains GPS coordinates of approximately 13600 taxis collected from August 3 to August 30, 2014, in the city of Chengdu. For each line, the data consists of the taxi ID, occupancy, time, longitude and latitude, with an average GPS sampling interval of 20 seconds.

By extracting continuous GPS records which was marked as occupied, the data of each trip will be identified from the three datasets. The time and coordinate of first and last GPS record marked as occupied will be set as the start and end time and coordinate of the trip respectively. These trip data will be used for further trip analysis and simulation.

### 4.1.2 Analysis of trip efficiency

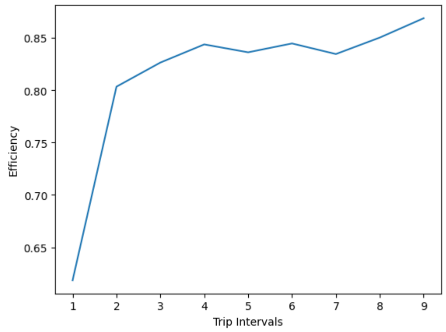
Before modelling and optimizing the pick-up and drop-off locations, it is necessary to simulate the characteristics of online ride-hailing trips based on past trip data. Therefore, in this subsection, the efficiency of the trip will be analysed.

The efficiency of a trip will be explained as the efficiency of the vehicle in reaching its destination during a specific segment of the trip. In the detour part, the trip efficiency is expected to be significantly lower than in the normal part of trip. To check the trip efficiency, the trips will be divided into 10 evenly distributed segments. Accordingly, the efficiency of each segment is defined as:

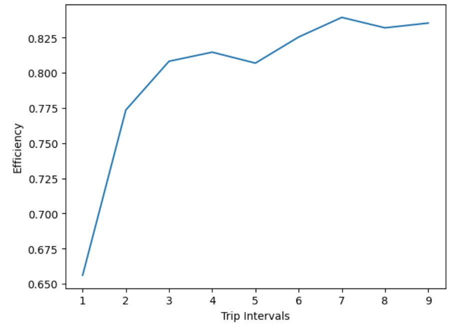
$$e_i = \frac{dist(P_{start,i}, D) - dist(P_{end,i}, D)}{\Delta s}$$

where $e_i$ denotes the efficiency of the segment $i$, $P_{start,i}$ and $P_{end,i}$ denote the start and end point of segment $i$, $D$ denotes destination of the trip, $dist(x, y)$ denotes the euclid distance between $x$ and $y$, $\Delta s$ denotes the distance this segment travelled.

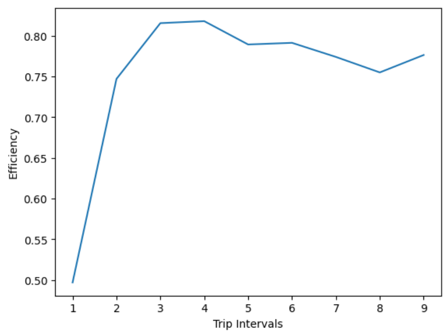
By analysing three datasets, the plot of trip segment efficiency can be shown as follow:



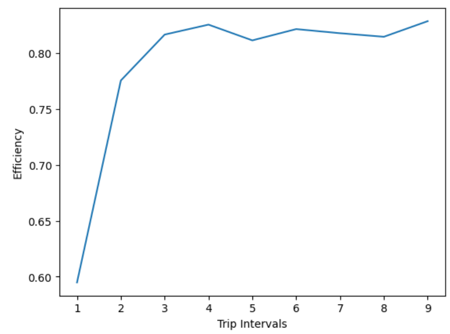

(a) New York Efficiency




(b) San Francisco Efficiency




(c) Chengdu Efficiency

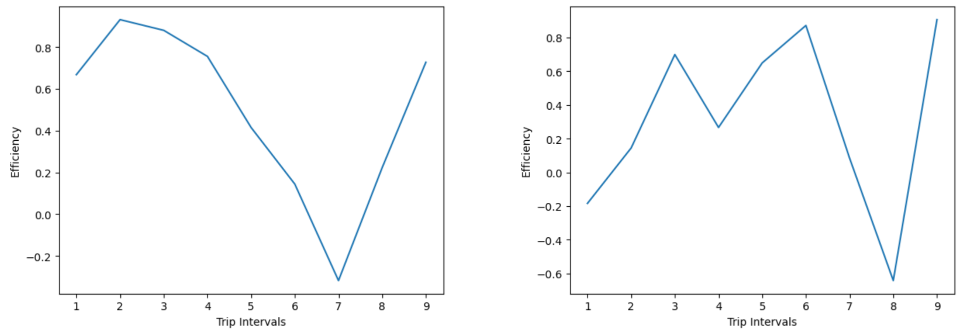



(d) Overall Efficiency

Figure 13: Trip Segment Efficiency

It can be seen that, averagely, the beginning part of the trip is significantly lower than rest part of the trip. And in some areas, the ending part of the trip was also below the average for the entire trip. This might indicate that the detours are more likely to appear at the beginning part of the trip. For some specific regions or trips, detours may also occur at the ending part.

Also, some trips with obvious detours were also selected and their efficiency graphs were plotted as follow:




(a) Detour appears one side

(b) Detours appear both sides

Figure 14: Trip examples with obvious detour

It can be found that the efficiency of the part where detour appears is usually less than 0, with the efficiency of the part nearby is less than 0.3. And this standard will be utilized in the following part of detour detection.

## 4.2 Assumption and Modelling

In this study, Python will be used to finish algorithms and data processing and analysis. Python has many prewritten libraries and packages that can be utilized to implement this project, which provide lots of functions required including downloading and converting maps, providing a priority queue solution, etc., which will reduce unnecessary workload and avoid duplication of work. In order to model the OpenStreetMap data as graphs that can be used in Python, in this study, *OSMnx* [6] will be used to download and model walking, driving or cycling street networks with lengths from OpenStreetMap, convert them into *MultiDiGraph* objects provided by *NetworkX* [25]. As a python package, *NetworkX* provides tools to create and manipulate graphs and contains many built-in routing methods. The street network will be stored as directed weighted graphs and its built-in shortest path method will be used as the standard solution of Dijkstra's and A* algorithms in this study. Additionally, in order to implement the priority queue required by the A* algorithm, *heapq* will be used to provide an implementation of the heap queue algorithm to create, push and pop elements [27].


(a) OpenStreetMap (b) Driving network

Figure 15: An example of converting OSM into MultiDiGraph object

## 4.3 Algorithm Design

In this study, in order to simplify the model, it is assumed that given a ride-hailing trip demand with a pair of origin and destination, its start point is located at the nearest street intersection. Additionally, before running the new algorithm, the shortest driving path will be found, and its length will be calculated so that the driving time can be used as the time limit and benchmark for the new algorithm. After the shortest driving path is found, it will be analysed to verified if there is a detour in the trip.

Based on the above analysis and assumptions, a new approach based on the bidirectional A* algorithm is designed to bypass potential driving detours and determine suitable pick-up and drop-off locations for online ride-hailing trips. The new algorithm aims to minimize the walking and driving distances while avoiding the detour, that is,

$$\min(dist_{walking}(O,p) + dist_{driving}(p,D))$$

where $dist_{walking}(x,y)$ and $dist_{driving}(x,y)$ denote actual walking and driving distance from $x$ to $y$, and $O$, $p$, $D$ denote *origin*, *pick-up point* and *destination* correspondingly.

Specifically, the algorithm consists of two parts: walking searching and driving searching.

Similar to bidirectional A* algorithm, this algorithm also maintains two priority queues to sort the points explored by searching boundary. Driving searching will be performed first from the side opposite to the detour as driving networks are usually simpler than walking networks, thus can save computational power. Another reason for this is that when observing the searching process of the greedy algorithm and A* algorithm, the search boundary will keep ”hit the wall” when it encounters the obstacle. In driving searching, the approximate location of the detour can be determined by the trip efficiency analysis performed in advance. After the driving searching reaches the obstacle, it will start recording the search boundary points with the best priority and record these coordinates as temporary meeting points. Here, the priority of driving searching is set as:

$$p_d = dist(P_i, D) + est(P_i, O)$$

where $p_d$ denotes driving priority, $dist(x, y)$ and $est(x, y)$ denote actual distance and estimated distance from $x$ to $y$, and $O$, $D$, $P_i$ denote *origin*, *destination* and *searching boundary* at $i$ th loop.

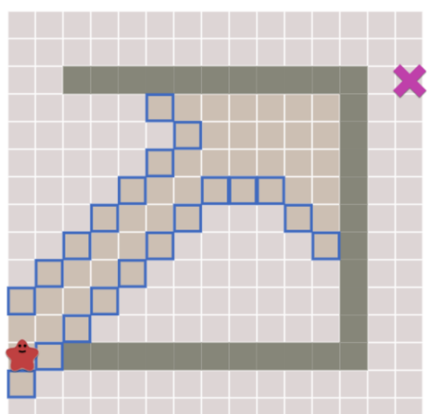


Figure 16: Search boundry of driving searching [4]

When driving searching approaches the detour and the first temporary meeting point being recorded, walking searching will begin, since the area around the obstacle needs to be sufficiently explored by driving searching, thus can avoid walking searching starting too early and missing the best potential meeting point. Unlike driving searching, walking searching will try to find the shortest path to the temporary meeting point given by driving searching, attempting to pass through the obstacles encountered in driving searching and explore possible walking shortcuts. The priority of the points in the search boundary is set as:

$$p_w = dist(O, P_i) + est(P_i, P_m)$$

where $p_w$ denotes walking priority, $dist(x, y)$ and $est(x, y)$ denote actual distance and estimated distance from $x$ to $y$, and $O$, $P_i$, $P_m$ denote *origin*, *searching bondary* at $i$ th loop, and *temporary meeting point* given by driving searching.

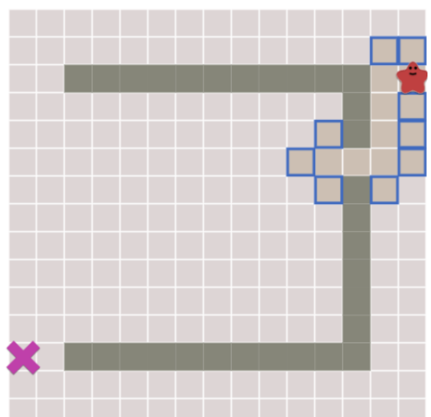
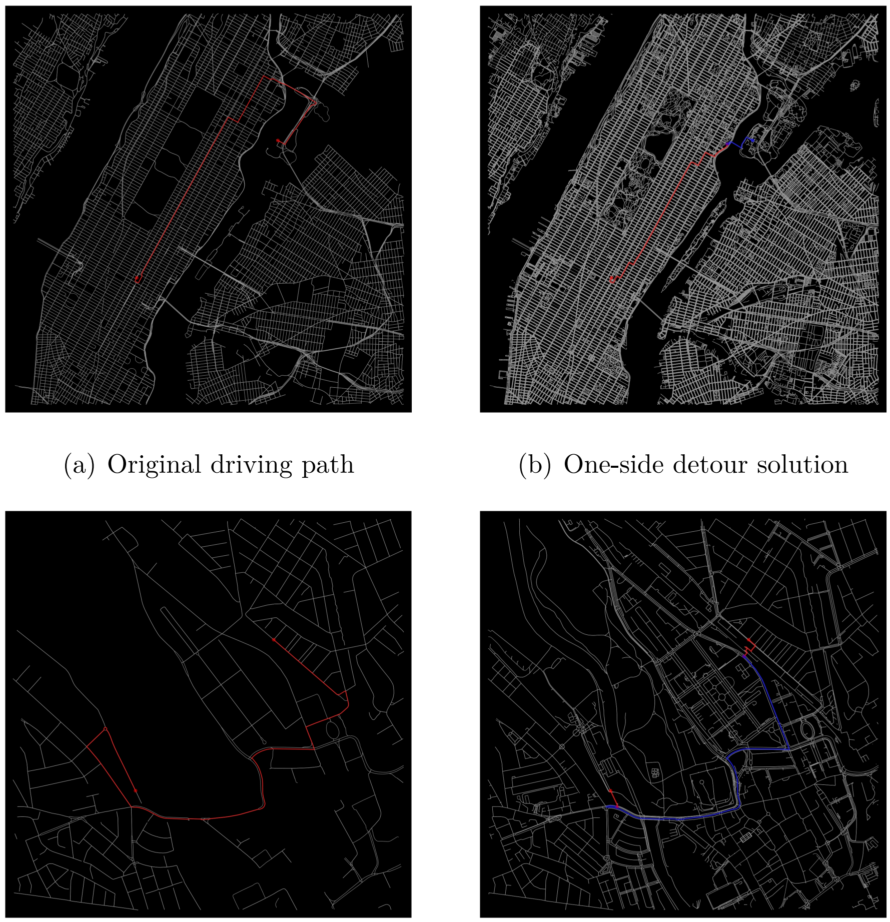


Figure 17: Walking searching might go through the driving obstacle [4]

When two searches intersect, the meeting point will be tested. Specifically, the meeting point must meet two standards: no detours in the driving path, and within time limit of trip. The purpose driving path detour checking is to ensure that the original detour has been avoided in given outcome. And the purpose of limiting the total trip time is to control the new total trip time within a certain range. Specifically, the new trip time should be less than the original trip time, that is, the time waiting for the car plus the time on the car, which can be defined as follow:

$$t_{limit} < \max(t_{walk}, t_{wait}) + t_{drive}$$

where $t_{limit}$ is driving time from *origin* to *destination* calculated before, $t_{walk}$ is walking time from *origin* to *meeting point* calculated by $t_{walk} = \frac{dist_{walk}}{v_{walk}}$, $t_{wait}$ denotes taxi waiting time, and $t_{drive}$ is driving time from *meeting point* to *destination* calculated by $t_{drive} = \frac{dist_{drive}}{v_{drive}}$.



(a) Original driving path (b) One-side detour solution

(c) Original driving path (d) Double side detour solution

Figure 18: Two examples for new algorithm

In summary, given a ride-hailing demand with origin and destination and the walking and driving networks, the driving searching will be run first. When the driving searching frontier has explored the driving obstacle, the walking searching will be started, attempting to cross the obstacle by walking. When the two searches meet, the meeting point will be tested. If the time limit is met and the detour has been bypassed, the walking path, the new pick-up point, and the new driving path will be determined. The general process is illustrated as follow:

**Algorithm 4:** New Algorithm

```
walk_frontier, drive_frontier ← PriorityQueue()
walk_frontier.put(start, 0); drive_frontier.put(destination, 0)
walk_came_from, drive_came_from, walk_cost, drive_cost ← {}
walk_cost[start], drive_cost[destination] ← 0
while walk_frontier.empty() or drive_frontier.empty() = False do
    if meet_point.empty() = False then
        walk_current ← walk_frontier.get()
        if walk_current_time < time_limit then
            break
        for next ∈ G.neighbors(walk_current) do
            new_cost ← walk_cost[walk_current] + G.cost(walk_current, next)
            if next ∉ walk_cost or new_cost < walk_cost[next] then
                walk_cost[next] ← new_cost
                walk_priority ← new_cost + heuristic(meet_point, next)
                walk_frontier.put(next, walk_priority)
                walk_came_from[next] ← walk_current
    drive_current ← drive_frontier.get()
    if drive_current_time < time_limit then
        break
    for next ∈ G.neighbors(drive_current) do
        new_cost ← drive_cost[drive_current] + G.cost(drive_current, next)
        if next ∉ drive_cost or new_cost < drive_cost[next] then
            drive_cost[next] ← new_cost
            drive_priority ← new_cost + heuristic(start, next)
            drive_frontier.put(next, drive_priority)
            drive_came_from[next] ← drive_current
            if next.getObstacle() and drive_priority < minimum_cost then
                meet_point = next
```

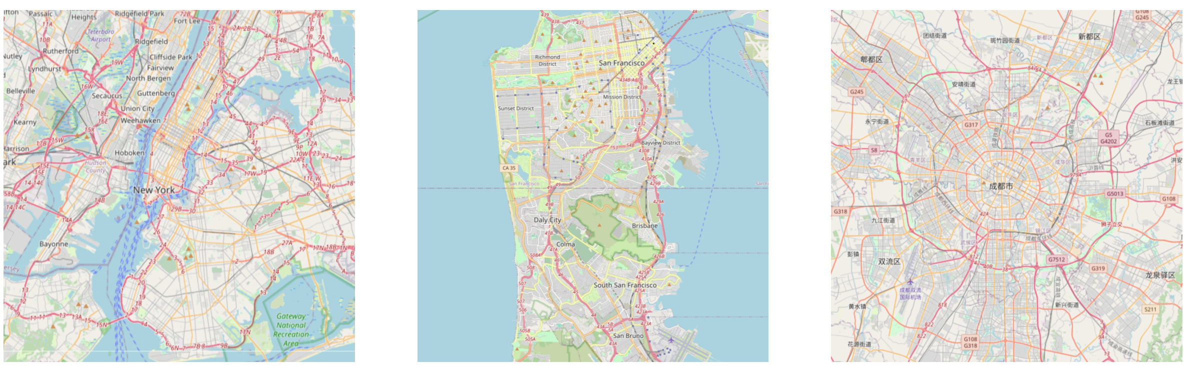
The above describes the running process of algorithm with one-side detour. For the case where there are detours on both sides, one side algorithm can be run firstly, and then run the algorithm between the other side and the newly given meeting point.

# 5 Numerical Simulation and Analysis

In this section, three real-world trip datasets from difference cities will be used to test the effectiveness and potential benefits of the new algorithm.

## 5.1 Simulation under Multiple Scenario

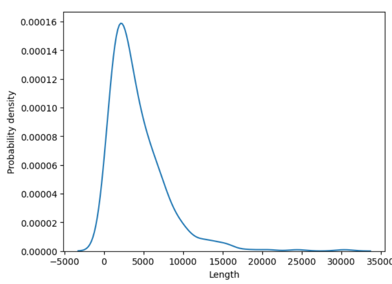
To quantify the benefits of integrating walking into ride-hailing routing, three datasets from New York, San Francisco, and Chengdu mentioned above will be utilized again.



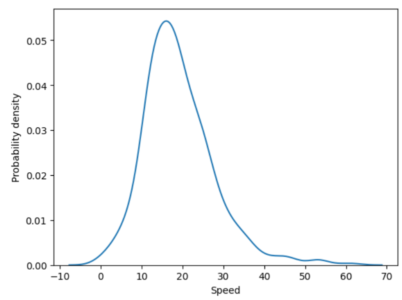
(a) New York (b) San Francisco (c) Chengdu

Figure 19: Map of three cities

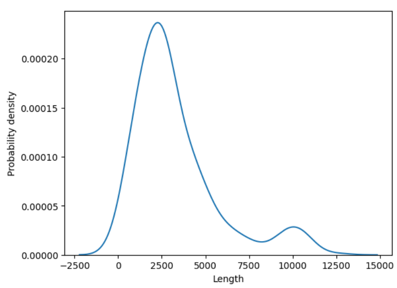
According to the map given by OpenStreetMap, it can be seen that the urban planning and structure of the three cities have different characteristics. Driving street network in New York city is divided by several large rivers, which means that New York is more likely to have geographical barriers. Also, Manhattan's grid road system contains a large number of one-way roads, which also increases the possibility of detours in ride-hailing trips. In comparison, San Francisco has a smaller population, smaller scale of street network and a grid road system similar to Manhattan in New York city. Chengdu has the largest area and the largest population among the three cities. Unlike the two cities in the United States, its road network has more directions and contains many expressways. These differences may cause these cities to have different variables required by the algorithm. Therefore, before analyzing the effectiveness of the algorithm, some variables required by the algorithm will be obtained from the analysis of three data sets, including length and speed of trips.



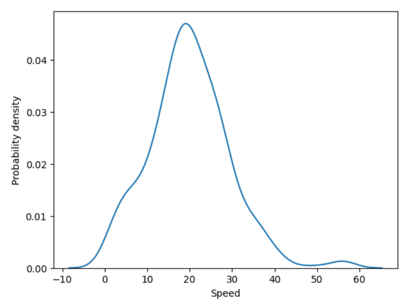

(a) New York trip length



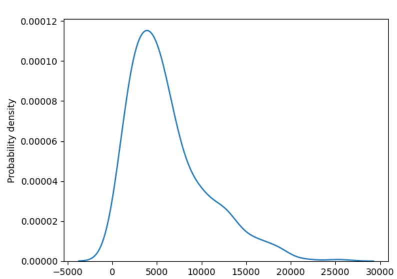

(b) New York trip speed



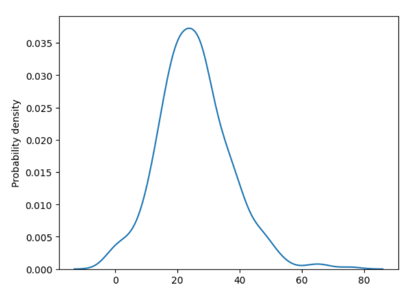

(c) San Francisco trip length




(d) San Francisco trip speed




(e) Chengdu trip length




(f) Chengdu trip speed

Figure 20: Distribution of Chengdu taxi trips

Through the analysis of the three datasets in figure 20, it can be found that the distribution of length and speed of trips in the three cities are slightly different. The average length of trips in New York city is 4257.3 metres, with an average speed of 19.20 kmph. And these numbers in San Francisco is 3296.3 metres and 19.95 kmph. A noteworthy detail is that there are two peaks in trip length around 10000 meters and trip speed at 55 kmph in the San Francisco graph, which may be caused by cross-city and airport trips. And Chengdu has the longest average trip length and the highest average speed among the three cities, at 6197.4 metres and 25.40 kmph. The different structures and different trip data of three different cities make sure that the use of three datasets can make the analysis more generalizable and representative. And in the following algorithm simulation analysis, the different speeds obtained from data analysis above will be used to fit the different conditions of different cities, making the analysis of the algorithm more accurate.

As the benefits of the algorithm related to user's preferences on walking, two strategies for constraining trip time including contains no extra time and contains extra 5 minutes will be tested. In terms of parameters, according to [22], the walking speed is assumed to be 5 kmph. And based on data from multiple studies [8] [61], the traditional ride-hailing service waiting time is assumed to be 5 minutes. The driving speed will be set to 20 kmph (New York city), 20 kmph (San Francisco), and 25 kmph (Chengdu) respectively based on the above analysis of the datasets.

## 5.2 Results, Analysis and Evaluation

The new algorithm was run to find new pick-up and drop-off points that bypass the detour under the three datasets. Respectively. the results are summarized in the following tables.

| Overall detour summary | San Francisco | New York | Chengdu |
|---|---|---|---|
| Percentage of trips with detour | 2.34% | 3.47% | 7.07% |
| % of double side detour in all trips with detours | 0.52% | 0.37% | 2.45% |

Table 1: Overall detour summary in three city datasets

Table 1 shows the overall detour situation of trips in three different cities. It can be seen that the proportion of trips with detours in different cities ranges from 2% to 7% of the total trips. Among them, the proportion of trips with detour in Chengdu is significantly higher than that in the other two American cities, which can be explained by the differences in urban planning and structures. Unlike the two cities in the United States, Chengdu has a relatively long history, which leads to different urban planning ideas between the new and old urban areas. In addition, there are many small rivers in Chengdu, together with many closed expressways and pedestrian streets. Also, because of the high population density in the urban area, there are many one-way roads in the city. All above phenomenon increase the complexity of the city street network, making the probability of detour in the urban area significantly higher than in the two American cities. This can be also proved by the higher probability of encountering trips with double side detours in Chengdu than that in the two American cities. But in general, trips with one side detour account for most of all detour trips, accounting for more than 97% in all the cities.

| Scenario: contains no extra time | San Francisco | New York | Chengdu |
|---|---|---|---|
| Reduced mileage for trips with detour | 24.44% | 24.33% | 30.19% |
| Total reduced VMT | 0.185% | 0.279% | 1.032% |
| Reduced time for trips with detour | 9.94% | 10.59% | 9.33% |
| Scenario: contains 5 mins extra time | San Francisco | New York | Chengdu |
| Reduced mileage for trips with detour | 23.29% | 23.34% | 29.04% |
| Total reduced VMT | 0.197% | 0.276% | 1.249% |
| Reduced time for trips with detour | 8.84% | 8.53% | 2.16% |

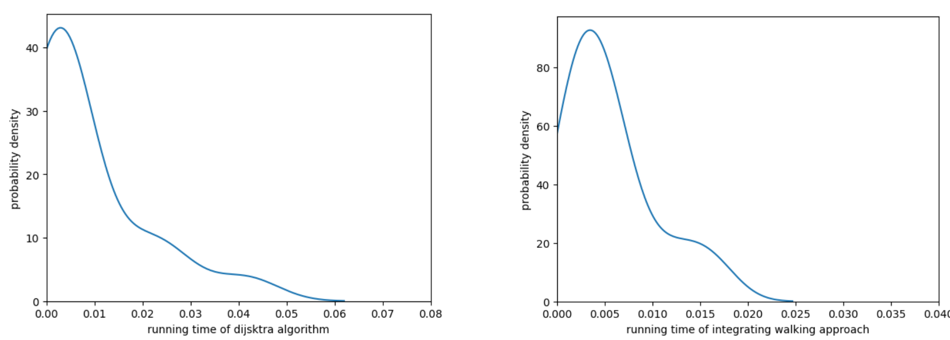
Table 2: Scenario simulation on potential benefits

Table 2 shows the potential benefit of integrating walking into routing, where VMT represents vehicle mileage travelled. The results reveal that ride-hailing efficiency can be improved when walking is allowed. For trips with detours, the new algorithm can reduce driving mileage by an average of 23% to 30% in various scenarios. It is notable that the reduction in mileage decreases slightly with more walking integrated, which may mean that the distance of walking is not proportional with its benefits. For the ride-hailing industry, the new pick-up and drop-off points can save the VMT of entire industry from 0.2% to 1.2%. Although the VMT reduction rate for all trips does not seem significant, it is the average of all trips, of which less than 10% of the trips can be optimized. In both aspects mentioned above, it is also notable that the numbers in Chengdu are higher than those in the two US cities, which may be the effect of urban planning and structure. In terms of the time reduction for the trip, the average time per trip is 2% to 10%. Among them, extending the walking distance will slightly reduce the amount of time reduction, which is in line with intuition.

| Scenarios | San Francisco | New York | Chengdu |
|---|---|---|---|
| % of failed pathfinding, without extra time | 5.11% | 11.11% | 18% |
| % of failed pathfinding, with 5 min extra | 7.88% | 9.74% | 6.77% |

Table 3: Percentage of failed pathfinding in different scenarios

In terms of performance of algorithm, as shown in table 3, in most cases, the algorithm can run stably, with the failure rate of pathfinding to below 10%. However, the complexity of urban planning and structure may affect the success rate, which is also where the algorithm needs to be improved in the next step.




(a) Running time of Dijkstra's algorithm

(b) Running time of new algorithm

Figure 21: Comparison between running time

At the same time, as shown in the figure 21, the new algorithm can give the above results within a reasonable running time. Compared with the widely used Dijkstra's algorithm, the new algorithm can complete it in 0.015 seconds in 95% of the cases, and for Dijkstra's, this number is 0.03 seconds. Also, compared with the traditional A* algorithm, the mean and standard deviation of the running time of the new algorithm have a slight increase, but at the same time it has achieved a bypassing of detour in the driving path.

In summary, three taxi datasets were used from three different cities for simulation experiments. The results show that in most cases, the algorithm provides a success rate of more than 90% within an acceptable running time. For the 2% to 7% of the total trips with detours, the average mileage can be reduced by 23% to 30% and the total travel time by about 2% to 10%. The mileage of the car-hailing industry can be reduced by 0.2% to 1.2%. In addition, the study also speculates that different urban structures and urban planning will have a significant impact on the detour rate, mileage reduction and other figures related to vehicle detours. However, it is worth noting that the above data is underestimated to a certain degree, which is caused by the taxi database is used as the data, and taxis imply the behaviours of passengers taking taxis on high-grade roads [19]. In addition, the successful rate of this algorithm under complex conditions may also reduce the benefits of this strategy.

# 6 Discussion

## 6.1 Conclusion

In traditional ride-hailing routing, the shortest driving distance from the pickup point to the drop-off point is usually prioritized. However, this may cause possible walking shortcuts to be ignored, and the hailed vehicles are forced to take detours.

Based on this fact, this study proposes a new pick-up and drop-off point optimization and routing algorithm based on the bidirectional A* algorithm, which utilizes the time passengers wait for ride-hailing vehicles to take a walk and takes the total travel time as a constraint to bypass the detour in the city. The algorithm is expected to reduce passengers' time in the car and fare without spending much cost and avoid online ride-hailing vehicles from entering congestion, one-way streets and unnecessary detours. For the online ride-hailing industry, it can improve vehicle operation efficiency, expand the potential users, reduce mileage and carbon emissions, enhance industry competitiveness and promote sustainable development of the industry.

To validate the effectiveness of the algorithm, this study uses taxi datasets in three cities: New York, San Francisco, and Chengdu for simulation experiments. The results show that in most cases, the algorithm provides a success rate of more than 90% within an acceptable running time. For the 2% to 7% of the total trips with detours, the average mileage can be reduced by 23% to 30% and the total travel time by about 2% to 10%. The mileage of the car-hailing industry can be reduced by 0.2% to 1.2%. In addition, the study also speculates that different urban structures and urban planning will have a significant impact on the detour rate, mileage reduction and other figures related to vehicle detours.

## 6.2 Reflection and Future Work

This study has several limitations, which should be addressed in future research. Firstly, in this study, three taxi datasets were used to simulate the trip characteristics of ride-hailing trips. Although online-hailing services are considered to have similar market demand, coverage area and trip length as traditional taxis [45], there are still some differences from traditional taxis and ride-hailing services [19], which makes the result and analysis of the algorithm inaccurate. In addition, this study has shown that the parameters of the algorithm for online-hailing vehicles detours are greatly affected by the region in section 5.2. This study is also limited to the datasets of three large cities. In the future, cooperation with online-hailing service providers should be reached to use the data of online-hailing service providers and verify the effectiveness of the new algorithm in actual operation. Also, in the processing of the dataset, it is assumed that users are picked up at the nearest intersection and the vehicles take the optimal route, which is not practical with the actual situation. This may lead to an incorrect estimation of the general trip and an underestimation of the benefit of the flexible pick-up and drop-off points given by the new algorithm.

On the other hand, the open-source map, OpenStreetMap, was used in this project. OpenStreetMap is very suitable as a tool for academic research as there are many packages and tools optimized for it. However, a regrettable fact is that the data provided by OpenStreetMap has many mistakes and inaccuracies. One notable example is that one-way roads for vehicles are also marked as one-way roads in the walking network. In addition, there are many intersections and bridges that are not updated in time, resulting in non-optimal routes in both traditional and new algorithm. In the future, cooperation with industry-leading and timely updated map and navigation service providers such as Google should be reached to ensure the accuracy of the road network.

In addition, regarding the design of the algorithm itself, the detection of detour is based on empirically formulated parameters. Now there are many more advanced tools such as AI and machine learning. The involving of these advanced tools is expected to achieve greatly improvements on accuracy and effectiveness of the algorithm. The marking and labeling of frequently used shortcuts and areas in advance can also provide a better user experience for this algorithm. In addition, the algorithm is not fully successful in finding pick-up and drop-off points, with the failure rate is around 9% mentioned in table 3. This might be caused by the inaccuracy of map or the design of the algorithm itself. It is expected that the algorithm can be improved in the future. Additionally, this algorithm gives a solution that avoids detour. Due to the usage of priority queue, it can be confirmed that the solution is acceptable within the constraints of time and detour detection. However, the algorithm has not yet been proven to give the optimal solution, which is considered a critical issue. In future work, proof of the optimality of the solution is necessary.

Finally, due to the lack of congestion data, the detour considered in this study is limited to obstacles in a geographical sense. If congestion data can be obtained, the conception of detour can be expanded. This algorithm may help more users and online car-hailing drivers.

In terms of personal experience and lessons learned in the research process, there is a lack of sufficient literature in this study. On the one hand, this means that there is a lack of sufficient data and previous research as a reference, and on the other hand, it means that this study has the potential to make a great contribution to this field. This paper innovatively creates a new algorithm based on bidirectional A*, filling the gap in previous research. The lack of ride-hailing trip datasets also has a negative impact on the progress of the project. This project had to use taxi datasets to simulate online car-hailing trip datasets. This is a major compromise for this project. In terms of personal skills, the author has broadened his understanding of the online car-hailing and transportation fields through this project. The learning and understanding of tools such as OSMnx has contributed to the author's continued development in this field. Learning such as ArcGIS, machine learning and AI in the future is expected to be beneficial to the development of this field. In terms of project and time management, due to the impact of illness, the timeline of this project was slightly delayed and was finally completed slightly later than expected. This reflects the shortcomings in project and time management.

# References


[1] Alemi, F. et al. (2018), What influences travelers to use Uber? Exploring the factors affecting the adoption of on-demand ride services in California *Travel Behaviour and Society*, Volume 13, pp.88-104, doi: https://doi.org/10.1016/j.tbs.2018.06.002

[2] Alemi, F. et al. (2019), What drives the use of ridehailing in California? Ordered probit models of the usage frequency of Uber and Lyft, *Transportation Research Part C: Emerging Technologies*, Volume 102, pp.233-248, doi: https://doi.org/10.1016/j.trc.2018.12.016

[3] Aliari, S. and Haghani, A. (2022), Alternative Pickup Locations in Taxi-sharing: A Feasibility Study, *Transportation Research Record: Journal of the Transportation Research Board*, 2677(1), pp.1391-1403, doi: https://doi.org/10.1177/03611981221104690

[4] Amit, P. (2014), Introduction to the A* Algorithm, Available at: https://www.redblobgames.com/pathfinding/a-star/introduction.html (Accessed: 14 Aug, 2024)

[5] Berger, T. et al. (2018), Drivers of disruption? Estimating the Uber effect, *European Economic Review*, Volume 110, pp.197-210, doi: https://doi.org/10.1016/j.euroecorev.2018.05.006

[6] Boeing, G. (2024), Modeling and Analyzing Urban Networks and Amenities with OSMnx Available at: https://geoffboeing.com/publications/osmnx-paper/ (accessed: 28 Aug, 2024)

[7] Brail, S. (2020), World cities of ride-hailing, *Urban Geography*, 43(1), pp.12-33, doi: https://doi.org/10.1080/02723638.2020.1775030

[8] Brown, A., and LaValle, W. (2021), Hailing a change: comparing taxi and ridehail

service quality in Los Angeles, *Transportation*, Volume 48, pp.1007-1031, doi: https://doi.org/10.1007/s11116-020-10086-z

[9] Cai, H. et al. (2019), Environmental Benefits of Taxi Ride Sharing in Beijing, *Energy*, Volume 174, pp.503-508, doi: https://doi.org/10.1016/j.energy.2019.02.166

[10] Cats, O. et al. (2022), Beyond the dichotomy: How ride-hailing competes with and complements public transport, *PLoS ONE*, 17(1), doi: https://doi.org/10.1371/journal.pone.0262496

[11] Chalermpong, S. et al. (2022), Ride-hailing applications in Southeast Asia: A literature review, *International Journal of Sustainable Transportation*, 17(3), pp.298-318. doi: https://doi.org/10.1080/15568318.2022.2032885

[12] Chen, J. et al. (2023), Improvement of an online ride-hailing system based on empirical GPS data, *Handbook of Mobility Data Mining*, Volume 3, pp.23-61, doi: https://doi.org/10.1016/B978-0-323-95892-9.00002-4

[13] CNBC (2018), Didi Chuxing, available at: https://www.cnbc.com/2018/05/22/didi-chuxing-2018-disruptor-50.html (accessed: 16 Aug, 2024)

[14] Contreras, S. D. et al. (2018), The effects of ride-hailing companies on the taxicab industry in Las Vegas, Nevada, *Transportation Research Part A: Policy and Practice*, Volume 115, pp.63-70, doi: https://doi.org/10.1016/j.tra.2017.11.008

[15] DataCastle (2016), GPS Data of Chengdu Taxi, Available at: https://js.dclab.run/v2/cmptDetail.html?id=175 (accessed: 29 Aug, 2024)

[16] Didi (2022), Annual Business and Social Responsibility Report, Available at: https://s28.q4cdn.com/896456191/files/doc_financials/2022/ar/2022-Annual-Business-and-Social-Responsibility-Report.pdf (accessed 27 Aug 2024)

[17] Diestel, R. (2017), *Graph Theory*, Springer, Heidelberger.

[18] Dijkstra, E. W. (1959), A Note on Two Problems in Connexion with Graph, *Numerische Mathematik*, Volume 1, pp.269-271, doi: https://doi.org/10.1007/BF01386390

[19] Ding, F., st al. (2024), Incorporating walking into ride-hailing: The potential benefits of flexible pick-up and drop-off, *Transportation Research Part D: Transport and Environment*, volume 127, article 104064, doi: https://doi.org/10.1016/j.trd.2024.104064

[20] Du, M. et al. (2020) Acceptance of electric ride-hailing under the new policy in Shenzhen, China: Influence factors from the driver's perspective, *Sustainable Cities and Society*, Volume 61, pp.102307, doi: https://doi.org/10.1016/j.scs.2020.102307

[21] Fielbaum, A. (2022), Optimizing a vehicle's route in an on-demand ridesharing system in which users might walk, *Journal of Intelligent Transportation Systems*, 26(4), pp. 432-447, doi: https://doi.org/10.1080/15472450.2021.1901225

[22] Fielbaum, A., et al. (2021), On-demand ridesharing with optimized pick-up and drop-off walking locations, *Transportation Research Part C: Emerging Technologies*, volume 126, article 103061, doi: https://doi.org/10.1016/j.trc.2021.103061

[23] Gomez, J. et al. (2021), Adoption and frequency of use of ride-hailing services in a European city: The case of Madrid, *Transportation Research Part C: Emerging Technologies*, Volume 131, pp.103359, doi: https://doi.org/10.1016/j.trc.2021.103359

[24] Grahn, R. et al. (2020), Socioeconomic and usage characteristics of transportation network company (TNC) riders, *Transportation*, Volume 47, pp.3047-3067, doi: https://doi.org/10.1007/s11116-019-09989-3

[25] Hagberg, A. A., et al. (2018), Exploring network structure, dynamics, and function using NetworkX, *Proceedings of the 7th Python in Science Conference (SciPy2008)*,

Gäel Varoquaux, Travis Vaught, and Jarrod Millman (Eds), Pasadena, California, USA, pp.11-15, Aug 2008

[26] Hall, J. D. et al. (2018), Is Uber a substitute or complement for public transit, *Journal of Urban Economics*, Volume 108, pp.36-50, doi: `https://doi.org/10.1016/j.jue.2018.09.003`

[27] Heapq - Heap Queue Algorithm (n.d.), Python 3.12.5 Documentation, Available at: `https://docs.python.org/3/library/heapq.html` (Accessed: 28 Aug, 2024)

[28] Henao, A. and Marshall, W. E.(2019), The Impact of Ride-hailing on Vehicle Miles Travelled, *Transportation*, Volume 46, pp.2173-2194, doi: `https://doi.org/10.1007/s11116-018-9923-2`

[29] Kaggle (2017), New York City Taxi Trip Duration, Available at: `https://www.kaggle.com/c/nyc-taxi-trip-duration/data` (accessed: 29 Aug, 2024)

[30] Kaszczyszyn, P. and Dutkowska, N. S.(2019), Walking Access to Public Transportation Stops for City Residents. A Comparison of Methods, *Sustainability*, 11(14), pp.3758, doi: `https://doi.org/10.3390/su11143758`

[31] Lazarus, J. R., et al. (2021), To Pool or Not to Pool? Understanding opportunities, challenges, and equity considerations to expanding the market for pooling, *Transportation Research Part A: Policy and Practice*, volume 148, pp. 199-222, doi: `https://doi.org/10.1016/j.tra.2020.10.007`

[32] Lee, K. et al. (2019), Impact of Ride-Hailing Services on Transportation Mode Choices: Evidence from Traffic and Transit Ridership, *Forthcoming at MIS Quarterly*, doi: `http://dx.doi.org/10.2139/ssrn.3244207`

[33] Li, X. et al. (2018), Modeling an Enhanced Ridesharing System with Meet Points and Time Windows, *PLoS ONE*, 13(5), doi: `https://doi.org/10.1371/journal.pone.0195927`

[34] Ma, N. et al. (2021) Balanced strategy based on environment and user benefit-oriented carpooling service mode for commuting trips, *Transportation*, Volume 48, pp.1241-1266, doi: https://doi.org/10.1007/s11116-020-10093-0

[35] Martinez, L. M. and Viegas, J. M. (2017), Assessing the impacts of deploying a shared self-driving urban mobility system: An agent-based model applied to the city of Lisbon, Portugal, *International Journal of Transportation Science and Technology*, 6(1), pp. 13-27, doi: https://doi.org/10.1016/j.ijtst.2017.05.005

[36] Megantara, T. R., te al.(2022), Strategies to Reduce Ride-Hailing Fuel Consumption Caused by Pick-Up Trips: A Mathematical Model under Uncertainty, *Sustainability*, 14(17), Article 10648, doi: https://doi.org/10.3390/su141710648

[37] Monteiro, C. M., et al. (2017), *Optimization of New Pick-up and Drop-off Points for Public Transportation*, Proceedings XVIII GEOINFO. Salvador, Brazil. 4-6 December. Available at: https://www.researchgate.net/publication/324482839_Optimization_of_New_Pick-up_and_Drop-off_Points_for_Public_Transportation (accessed: 20 Aug, 2024)

[38] Moran, M. (2021), Travel Decision Survey 2021, available at: https://www.sfmta.com/sites/default/files/reports-and-documents/2022/04/sfmta_td2021_rpt_v2.pdf (accessed: 16 Aug, 2024)

[39] Nie, Y. (2017), How can the taxi industry survive the tide of ridesourcing? Evidence from Shenzhen, China, *Transportation Research Part C: Emerging Technologies*, Volume 79, pp.242-256, doi: https://doi.org/10.1016/j.trc.2017.03.017

[40] Open Street Map (n.d.), *Copyright and Licence*, Available at: https://www.openstreetmap.org/copyright (accessed: 22 Aug, 2024)

[41] Open Street Map (2024), *Projection*, Available at: https://wiki.openstreetmap.org/wiki/Projection (accessed: 22 Aug, 2024)

[42] Peter, E. H. et al. (1968), A Formal Basis for the Heuristic Determination of Minimum Cost Paths, *IEEE Transactions on Systems Science and Cybernetics*, 4(2), pp.100-107, doi: `https://doi.org/10.1109/TSSC.1968.300136`

[43] Piorkowski, M. et al. (2009), CRAWDAD epfl/mobility, *IEEE Dataport*, doi: `https://dx.doi.org/10.15783/C7J010`

[44] Qiao, S. and Yeh, A. G. (2023), Is ride-hailing competing or complementing public transport? A perspective from affordability, *Transportation Research Part D: Transport and Environment*, Volume 114, pp.103533, doi: `https://doi.org/10.1016/j.trd.2022.103533`

[45] Rayle, L. et al. (2016), Just a better taxi? A survey-based comparison of taxis, transit, and ridesourcing services in San Francisco, *Transport Policy*, Volume 45, pp.168-178, doi: `https://doi.org/10.1016/j.tranpol.2015.10.004`

[46] Roder, C. (2018), The Effects of Ride Hailing Services on Travel and Associated Greenhouse Gas Emissions, *UC Davis: National Center for Sustainable Transportation*, Available at: `https://escholarship.org/uc/item/2rv570tt` (accessed: 26 Aug, 2024)

[47] Santi, P. et al. (2014), Quantifying the Benefits of Vehicle Pooling with Shareability Networks, *Proceedings of the National Academy of Sciences*, 111(37), pp.13290-13294, doi: `https://doi.org/10.1073/pnas.1403657111`

[48] Sarma, N. et al. (2023), On-Demand Ride-Pooling with Walking Legs: Decomposition Approach for Dynamic Matching and Virtual Stops Selection, *SSRN* doi: `http://dx.doi.org/10.2139/ssrn.4383040`

[49] Slowik, P. et al. (2019), Assessing ride-hailing company commitments to electrification, *The International Council on Clean Transportation*, Available at: `https://theicct.org/wp-content/uploads/2021/06/EV_Ridehailing_Commitment_20190220.pdf` (accessed: 21 Aug, 2024)

[50] Slowik, P. (2019), Why aren't Uber and Lyft all-electric already?, *The International Council on Clean Transportation*, Available at: https://theicct.org/why-arent-uber-and-lyft-all-electric-already/ (accessed: 21 Aug, 2024)

[51] Tang, B. et al. (2019), How app-based ride-hailing services influence travel behavior: An empirical study from China. *International Journal of Sustainable Transportation*, 14(7), pp.554-568. doi: https://doi.org/10.1080/15568318.2019.1584932

[52] Tian, G. et al. (2023), Exploring the influences of ride-hailing services on VMT and transit usage - Evidence from California *Journal of Transport Geography*, Volume 110, pp.103644, doi: https://doi.org/10.1016/j.jtrangeo.2023.103644

[53] Tirachini, A. and Rio, M. D. (2019), Ride-hailing in Santiago de Chile: Users' characterisation and effects on travel behaviour, *Transport Policy*, Volume 82, pp.46-57, doi: https://doi.org/10.1016/j.tranpol.2019.07.008

[54] Tirachini, A. (2020), Ride-hailing, travel behaviour and sustainable mobility: an international review, *Transportation*, Volume 47, pp.2011-2047, doi: https://doi.org/10.1007/s11116-019-10070-2

[55] Uber (2023), Uber Announces Results for Fourth Quarter and Full Year 2023, Available at: https://investor.uber.com/news-events/news/press-release-details/2024/Uber-Announces-Results-for-Fourth-Quarter-and-Full-Year-2023/default.aspx

[56] Wenzel, T. et al. (2019), Travel and energy implications of ridesourcing service in Austin, Texas, *Transportation Research Part D: Transport and Environment*, Volume 70, pp.18-34, doi: https://doi.org/10.1016/j.trd.2019.03.005

[57] Wikipedia (n.d.), Directed graph, Available at: https://en.wikipedia.org/wiki/Directed_graph

[58] Wong, R. C. P. et al. (2014), Bi-level decisions of vacant taxi drivers traveling towards taxi stands in customer-search: Modeling methodology and policy implications, *Transport Policy*, Volume 33, pp.73-81, doi: https://doi.org/10.1016/j.tranpol.2014.02.011

[59] Wu, T. et al. (2018), Development and application of an energy use and CO2 emissions reduction evaluation model for China's online car hailing services, *Energy*, Volume 154, pp.298-307, doi: https://doi.org/10.1016/j.energy.2018.04.130

[60] Yan, L. et al. (2020), Quantifying and analyzing traffic emission reductions from ridesharing: A case study of Shanghai, *Transportation Research Part D: Transport and Environment*, Volume 89, pp.102629, doi: https://doi.org/10.1016/j.trd.2020.102629

[61] Yang, H., te al. (2021), Equitable? Exploring ridesourcing waiting time and its determinants, *Transportation Research Part D: Transport and Environment*, volume 93, article 102774, doi: https://doi.org/10.1016/j.trd.2021.102774

[62] Zhang, Y. and Zhang, Y. (2018), Exploring the Relationship between Ridesharing and Public Transit Use in the United States, *International Journal of Environmental Research and Public Health*, 15(8), article 1763, doi: https://doi.org/10.3390/ijerph15081763

[63] Zhu, P and Mo, H. (2022), The Potential of Ride-pooling in VKT Reduction and Its Environmental Implications, *Transportation Research Part D: Transport and Environment*, Volume 103, pp.103155, doi: https://doi.org/10.1016/j.trd.2021.103155