Design a Customized Bus-Booking Platform

Project for Algorithm and Complexity

Xiaofeng Gao
Department of Computer Science,
Shanghai Jiao Tong University, Shanghai, China

Abstract. This project introduces the concept of a customized bus-booking platform. It includes background, brief introduction, and advantages of such platform. As consultants, you are required to complete the kernel design of such new platform, including route planning, order dispatch, and pricing strategy, etc. Please read this document carefully and complete the corresponding tasks.
Keywords: Bus-Booking, Order Dispatch, Route Planning, Pricing Strategy

1 Bus-Booking Platform

With the rapid expansion of urban scale and population growth, the demand for urban transportation within city grows rapidly in recent years. Traditional public transportation, such as buses, subways, and taxis, cannot fully meet people's demand for the timeliness and convenience of urban traffic. To this end, some novel transportation concepts have been proposed and put into the wide application, such as bike-sharing and car-hailing. They provide more choices for people and effectively relieve the increasingly pressure of urban traffic.

However, all existing methods still cannot completely satisfy the needs from passengers, especially for some locations at special periods. E.g., according to the official statistics from Shanghai Traffic Management Bureau (STMB), the number of the passengers arrived at Shanghai Pudong International Airport is about 18,000 during midnight period (23:00-02:00). However, the local public transport capacity can only provide pick-up service for around 8,200 passengers, including 7,000 passengers by taxi and online car-hailing, and 1,200 passengers by airport shuttle buses (less than 7% of the total arrived passengers), leaving a large number of passengers waiting at the airport for a long duration.

In order to provide higher-quality service, STMB intends to develop a new service APP named bus-booking platform. Unlike car-hailing platforms, a bus-booking platform can dispatch several passenger orders to one bus together, compute a specially planned route cycle based on these requirements, and then ride the passengers to their destinations sequentially. Compared to car-hailing, bus-booking can effectively utilize vehicle resources, reduce energy consumption, and save road resources. Simultaneously, passengers could be served quickly with much lower price. In all, combining the benefits of bus service and car-hailing, this new APP provides economical service with real-time flexible route planning and order dispatch, which contributes to a better user experience.

As consultants, your team is required to complete the kernel design of such new platform, including destination selection, route planning, order dispatch, and pricing strategy, etc. Please complete the following tasks respectively.

2 Design Task

In order to build such a platform, we need four steps: destination selection, order dispatch, route planning, and pricing tickets. Firstly, there are too many and flexible destinations for passengers, and thus we need to identify a set of hot destinations as our "stations" (denoted as D), such that passengers will choose one of them as their destinations, instead of pointing an arbitrary location. Next, you need to schedule passengers to different buses and plan the route for each bus, so that we can use minimum energy consumption and vehicle resource to earn maximum profit. Note that the bus route is a cycle that starts from a departure station and returns back to this station again. Finally, for the sake of profits, you need to set the bus fare rules and explain the rationality.

2.1 Destination Selection

In the early stage of this service, it is complicated to compute arbitrary passenger destinations to form a route. Instead, STMB hopes to select k (k could be 30, 50, etc.) hot destinations, such that passengers can select one of them as their destination station. Please help STMB to choose these hot stations according to various factors such as position, district, operation cost, and historical records from taxis, subways and buses, etc.

2.2 Problem Formulation

Suppose that there are n passengers and m buses at the departure station, while the destination station of passenger i is $d_i \in D$ and the capacity limitation of each bus is L. The bus fuel cost per kilometer is c_b. The cost per route of a bus is c_r (You can consider it as a fixed management fee, like driver remuneration and administration expenses, to complete a lap). The bus fare price for one passenger from the departure station to station d_j is p_j. STMB requires you to dispatch passengers to different buses and plan routes for each bus, in order to earn maximum profits. (Attention: each bus may have distinct route, and every passenger starts at a unique departure.)

There are many possible ways to formulate the *order dispatch and route planning* problem (ODRP), some of which are illustrated as follows.

1. Suppose STMB wants to satisfy every passenger (this implies $n \leq mL$), you need to complete the task of order dispatch and plan the route for each bus to maximize the total profit (ODRP-V1).
2. If STMB does not satisfy all passengers (either $n > mL$ or you can ignore some orders). You need to pick a subset of orders, allocate them to different buses and plan the route for each bus to maximize the total profit (ODRP-V2).
3. Dropping orders definitely brings dissatisfaction and loses potential customers. Assume $n \leq mL$ again (while the carrying capacity of STMB is sufficient, taking more passengers may instead decrease the profit). Now your goal is to maximize the total profit and minimize the user dissatisfaction simutanously (ODRP-V3). Please give quantitative definition of "user dissatisfaction" appropriately.

Give formal definitions of these three ODRP problems. Do you have any other considerations of this problem? Please try to provide the 4th version of ODRP problem yourself. You are free to add additional variables and descriptions that are omitted or missed in the previous statements.

2.3 Theoretical Analysis

Before designing algorithms, You need to distinguish the difficulity of these ODRP problems and conduct deep theoretical analysis.

1. Formulate all ODRP problems as programmings.
2. Give the decision versions of these problems.
3. Consider the difficulty of these ODRP problems. Are they P? NP? NP-Complete? Or different version belongs to different category? Prove with your justifications.

2.4 Algorithm Design

Let us pick ODRP-V2 as an example to design algorithms. For small scale cases, try to get an optimal solution effectively through any standard programming solver (or whatever you want to use to find an optimal solution). For large scale cases, design an algorithm for feasible solutions, prove your correctness and distinguish the time complexity. As m, n, D increases, when will your computer's computing ability reach the physical limit to compute an optimal solution? Try to explain theoretically or practically.

2.5 Pricing Strategy

Eventually, the goal of STMB is to make profits. Please design a reasonable pricing mechanism for this bus-booking platform according to traditional bus fare rules, taxi price lists, and user psychology, etc. You need to explain the rationality of your pricing strategy.

3 Numerical Experiments

Now let us use real-world trace data from DIDI GAIA Open Dataset* to test the efficiency and effectiveness of your design. We have two order history sets for taxis, dripping express and special cars, described as follows:

- The first one is the order history from Haikou city during May 1st, 2017 to October 31st, 2017. It covers 184 days with 14,159,502 piece of records. Each student group will assign 4 to 5 day period for computation, roughly 330k piece of records per group.
- The second one is the order history from Chengdu city during November 1st to November 30th, 2016. It covers 30 days with 7,065,937 piece of records. Each student group will assign 1 day period for computation, roughly 240k piece of records per group.
- Data downloading link: `https://jbox.sjtu.edu.cn/l/dn1iEl`

The features of each history data set are introduced in Project-DataField.pdf, including descriptions of field type, sample, and comment.

You are required to complete the following experiments respectively (You need to conduct experiments separately based on two sets of data provided). Note that each group will assign a special period for each data set, which will greatly reduce the workload. The detailed period assignment for each group is listed in Project-DataAssignment.pdf.

3.1 Destination Selection

Select k hot destination stations (set $k = 30$ and $k = 50$ in your report, and will be set arbitrarily via Demo) according to your destination selection strategy in Sec. 2.1. If the strategy you proposed cannot be implemented because of the restriction of dataset, you need to adjust your strategy to suit the current dataset.

3.2 Departure Station Selection

Note that we are dealing with passenger transportation problem at specific locations for burst traffic, usually airports or railway stations, etc. Please investigate the data set in your hand carefully and choose the most popular departure location as a "virtual airport". Since passengers may call car-hailing service anywhere in a city, you need to merge arbitrary locations as a big "metanode" with radius of r km (r is usually set between 3 km and 10 km). You can adjust this parameter appropriately to get a best result.

3.3 Simulations

Filter the orders from your selected departure station, and pick 30 hot destination stations AGAIN based on these selected orders (according to your adjusted strategy in Sec. 3.1). Assume $L = 20$, 25, 30, 35, 40, 45 respectively (you need to consider at least three of them), you need to determine different m values according to the orders and L, and compare your designs in Sec. 2.4 (optimal vs feasible solutions under different settings). Visualize your results and profits (using pricing strategy in Sec. 2.5) clearly on different data sets.

Please consider eventually based on your data and results: what are the best settings for L and m, if all other parameters are known?

3.4 Comparison with Taxis

You can find the taxi price, fuel cost, and driver remuneration etc. for each city through the Internet. Based on results in Sec. 3.3, compare Taxi System with your bus-booking platform (assume all passengers are taking taxis, or taking your buses). Adjust and illustrate the rationality of your pricing strategy.

*https://outreach.didichuxing.com/research/opendata/en/

4 Report Requirements

You need to submit a report for this project, with the following requirements:

4.1 Writing Requirements

1. Your report should have the title, the author names, IDs, email addresses, the page header, the page numbers, figure for your simulations, tables for discussions and comparisons, with the corresponding figure titles and table titles.

2. Your report is English only, with a clear structure, divided by sections, and may contain organizational architecture like itemizations, definitions, or theorems and proofs.

3. Please include reference section and acknowledgement section. You may also include your feelings, suggestion, and comments in the acknowledgement section.

4. Please define your variables clearly. If needed, a symbol table is strongly recommended to help readers catch your design.

5. Please visualize your report appropriately such that readers could understand your design easily, and see your results directly.

6. Please provide the information you collected such as taxi price, fuel cost, driver remuneration and cite their source.

4.2 Uploading Requirements

1. Please include your latex source and simulation codes upon submission. Please also include the code for visualization, e.g., route visualization, grouping visualization, etc.

2. Please include the latitude and longitude data list of the destination stations selected in Sec. 3.1.

3. Please include the filtered order records based on your departure station selected in Sec. 3.2 (should be two lists from two history records).

4. Similarly, include the latitude and longitude data list of the departure station selected in Sec. 3.3 and the corresponding destination stations. Calculate and submit the distance matrix between stations (should be realistic distance on city road network).