Analysis and Visualization of Online Ride-Hailing Order Data

Project for Principles and Practice of Problem Solving

1 Introduction

Recently, the popularity of online ride-hailing systems, such as Didi and Uber, brings significant convenience to daily transportation. However, they are all facing notorious challenges, such as demand-supply gap. Suppose you are a data analyst, and are required to analyze and visulize order-related data to help the company solve challenges.

2 Dataset Description

The dataset includes 15 days order data in Chengdu city, partially from KDD Cup 2020. Here are some useful fields:

Table 1. Samples of Chengdu Order dataset.

fields	type	description	example
order_id	String	ID of order	eb9dd4095d9850e6287cefd813775a6c
depature_time	Float	Ride Start Time	1477964797
end_time	Float	Ride End Time	1477966507
orig_lng	Float	Pick-up Longitude	104.092097319258
orig_lat	Float	Pick-up Latitude	30.7063482074305
dest_lng	Float	Drop-off Longitude	104.08673269432
dest_lat	Float	Drop-off Latitude	30.653252740962
fee	Float	Fee of order	3.54

In order to meet the requirements of refinement, we divide the main urban areas of Chengdu into 10×10 grids, as shown in the figure below. The latitude and longitude of each lattice has been given. The program also needs to be able to analyze and visualize the order information in the small grids.

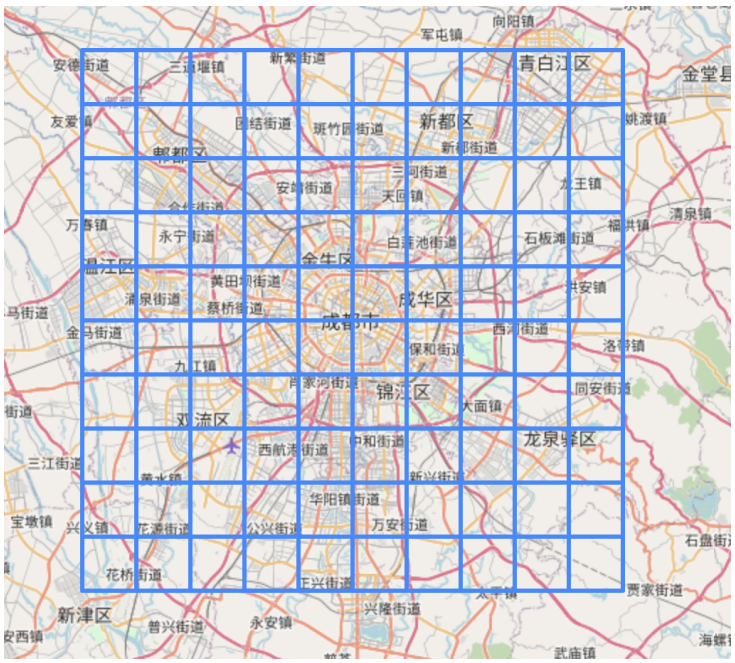

Fig. 1. Grid illustrated in Chengdu city.

3 Tasks

3.1 Mandatory Part

Design a graphic user interface, which can

1) load the dataset. Allow users to select the **fields** and **time span** of users' interest. A progress bar is required to show the loading progress.
2) display the spatio-temporal demand patterns. For example, **the number of orders** in Chengdu over time, the number of orders in each fine-grained grids over time. Display forms include but not limited to the line chart, histogram, pie chart, etc. Users should be able to **tune the parameters**, such as the starting time, ending time, time step and grid id, graphically. **Interpolation** is required to smooth the line when only a few data points are selected.
3) display other valuable information. For example, the distribution of users' **travel time, order fees**, etc. Moreover, total revenue of the company in a certain period of time is also displayable.

Hint: Codes need to be implemented via multi-thread to prevent your program from being stuck when you are computing while displaying.

3.2 Elective Tasks

Implement additional analysis by yourself. Use your knowledge and freely exert your creativity, and we won't limit views from which data is analyzed and displayed. **Please note that one can implement up to two functions in elective tasks. Functions that are more than two will not be counted in the total score**. Here are some recommendations:

1) Map-related analysis.
   - Show the spatial distribution of order origins on the map, such as the thermal diagram in a period of time;
   - Show how people transfer, i.e., the connections from the origins to the destinations;
   - Route planning: input starting and ending positions, and display the recommended route on the map.

   Hint: We recommend openstreetmap library.
2) Prediction-related analysis.
   - Given the origin and the destination, how long will it take for a passenger to arrive?
   - Given a starting position, where may passengers go?
3) Query-related analysis.
   - Retrieve orders: return a set of similar orders based on the input order information.

4 Submission Requirements

The project will be graded on the basis of your **project report**, **video demo** and the **source codes**. You need to zip and upload these files to the Canvas. It is due by **11:59pm on December 27th, 2020**. **No late submission will be accepted.**

4.1 Report Requirements

1. No more than 5 pages. No formatting requirements.
2. Should at least contain
   a) **Introduction.** Describe your main ideas and what you do in this project.
   b) **Implementation Details.**
   c) **Results.**
   d) **Discussions.** You could discuss the performance of your application or discuss the interesting results you have revealed from the data.
3. Report is in English only.

4.2 Demo Requirements

1. Make a video to present your work by screen recording or camera shooting. Please check that the video is playable!
2. No more than 5 minutes.
3. Presentation is in English only.

Examples of screen recording:

- YouTube Demo
- Bilibili Demo

Examples of camera shooting:

- Presentation Demo

4.3 Source Code Requirements

1. Add a README file, which includes instructions to run your codes.
2. Make sure it runs smoothly on your own computer.