Independent Study Research Final Report

Name:Dengpan Yuan

1 Introduction

In this semester I had the independent research with my instructor Yu Yang, we generally had two projects in total. We did the research around the indoor positioning system for food delivery path planning and the vehicle mobility for the city planning. In the first project - Translock:Transparent Indoor Localization with Uncertain Anchor Information for Instant Delivery, I basically did the work on the Ground Truth Data Processing and Find the Retailer Map from Baidu Map. And for the second project Vehicle Mobility - Modeling human exploration behaviors in Physical World, I did the work on Detecting Hub and Processing the Data for Generating the graph. In the next section, these two projects will be illustrated precisely.

2 Research Project

2.1 TransLoc

This research is under the background that instant delivery is an important urban service in recent years driven by the increasing consumer for quick product delivery, which generally requires to be delivered within half an hour in an urban area. One important issue for the business is to keep updating the status of carries especially the real-time locations, which is challenging when they are in an indoor environment. Therefore, in this research, our aim is to address two challenges including uncertain anchor reporting behaviors and indoor mobility behaviors. The main contribution from me is:

- Processing the Ground Truth of Carriers' Indoor Locations
- Summarizing Ten City Data in China including Area and Population
- Positioning the Retailer's Map from Baidu Map

2.1.1 Process the Ground Truth Data

In this task, I was given a few files including carriers' indoor location and beacon id, my instructor wants me to find the path for every single carrier man from check_in_record_bailian.csv to map the signal_bailian.csv. Besides, there are a few other non-critical tasks including:

- Add RSSI(Received Signal Strength Indictor) Columns
- For Each tracking_id, for Each passing_beacon_id, get MAX/MIN RSSI value.
- Draw Box Plot, Based on MINMAX Distribution for the RSSI MAX/MIN value we got before.
- Get Interval Time of Each Passing becaon_id

The basic method I adopted was using python csvLib to process these two csv files, since this is the most familiar way for me to deal with data and also it's a good implementation given not converting files to SQL or other file types. What I processed was loading two files into memory and adding internal conditions: when arrive_rst_at_iot != "" and the date is 2019-05-08, plus the time is over 30s then split. Although it's pretty easy to write the SQL codes to group the data according to tracking_id but the internal conditions will make it harder to intercept the variables with linear looping flow. So based on my consideration, the smart way may not be the most efficient way because it might consume more time on debugging, so I choosed looping the data linearily and setting the conditions. Finally for the RISS value it's pretty easy to find the MIN&MAX from my created dict data structure. Then drawing the Box Plot I used seaborn blox plot module to make it possible, it's also an efficient and neat way to visualize the data.

2.1.2 Summarizing City Data

This task requires me to summarize Area and Population data from ten cities: Shenzhen, Guangzhou, Beijing, Tianjing, Shanghai, Hangzhou, Wuhan, Haerbin, Hefei, Xi'an. I basically googled from wikipedia and get these data. To ensure the accuracy, I also tried city summary in Baidu and certain the final data after comparison.

2.2 Positioning the Retailer's Map from Baidu Map

In this task my instructor gave me 30 around restaurants in a mall in Shanghai and asked me to position them in different floors where separated in 5 floors in total. The most challenge is that the map of the mall is so big that I need to edit and merge them with a huge figure for every single of 5. This requires a little bit editing skill and some time, so finally I finish the task with a nice joint. The following is a show map:

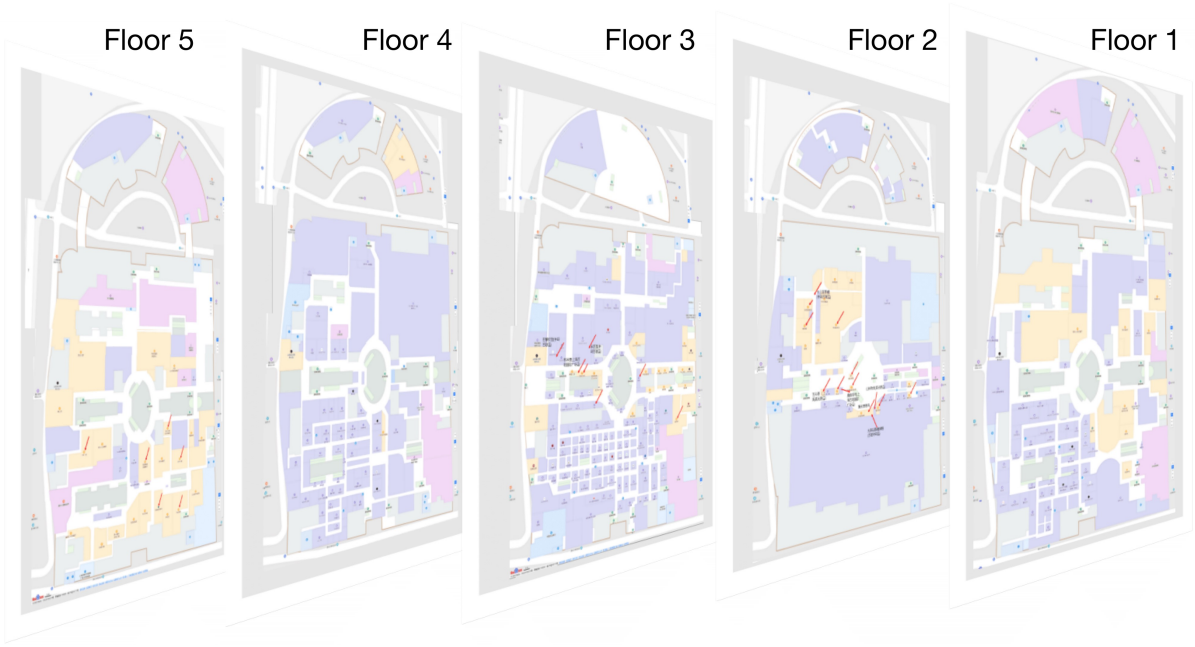

Figure 1: Floor Map

2.3 Vehicle Mobility

The background of this project is that given the well study of Human mobility, the applications of this is wide applied, such as urban plan, disease control and social networks. Several studies have shown that human mobility typically follows regular patterns. For example, in urban areas, people typically speed most of the time around a few major hubs such as home and work locations and periodically commute between them. Furthermore, people frequently visit a few minor hubs around or between their major hubs, such as shopping markets, restaurants and gyms. Therefore, in this project I was mainly responsible for detecting hubs based on designed algorithm.

2.3.1 Detecting Hubs

In this project my instructor send me 6 files about 100m_source_dest data, they contain records about user for the trip, and locaiton(longtitude and latitude) data, the most important data is the centroid_id, which is formed with, several grids, each of trip will fill in the centroid grid based on their location. Therefore, my task is basically following:

- For Each User, Group(center_row_id, center_col_id) and count them
- Sort the Data According to The City Index
- Predict the Probability of for each trip
- For Each User, We Have a Probability List
- Retrive the Interval Time and Split if it's over 30

The procedure I did for the group is still using Python csvLib package, because it's a linear looping with internal conditions, therefore, python takes the advantage again here. But in this time, it has a difference for the conditions I met in the previous project - that is I have to count the city meet times which is used in the final probability statistics. My final is creating a dict again for the each trip and do a group by for the grouped trip data again to count the times. The second part is counting the occruence time and get the probability of a certain trip in the city over all of these records, but for never reached city I need to pad 0 to avoid the divides 0 error. Except this, I think the most challenging part is splitting the trip based on 30 seconds interval time limit, because it needs you to always keep a pointer for every trip and a map list to record the which city arrived and most recent leave time, in this way could you get the time difference and judge based on splitted time, if there's over 30 seconds, then just created another record with different trip_id for this splitted time. The final part is grouping the probability for different cities, this is also easy to achieve by looping linearily within O(n) time space.

3 Conclusion

This independent research is the first time I researched on the field of smart city. From the experience of these two projects, I found it requires a heavy data processing skill, and also the reliable data source is critical for conducting the research, data visualization is also necessary for presenting our research result persuading and directly. And also a good communication with the instructor is important for moving the research which we discussed a lot during the semester. So here I want to thanks for the guidance from my instructor Yu Yang and Prof.Zhang in the semester!