Project for Principles and Practice of Problem Solving

Gowalla POI Visualization System

Weiming Zhang

January 1, 2022

1 Introduction

Gowalla is a location-based social networking website where users share their locations by checking-in. The friendship network is undirected and was collected using their public API. The data provided by Gowalla is presented in the form of Point-of-interest. In this project, the author implemented a POI data visualization system, in which users can learn about the spatio-temporal distribution of the social network. In this project, users can also analyze the distribution pattern and trend, make comparisons of different users and POIs, and make predictions on a given user or POI.

2 Implementation Details

2.1 Dataset Import

The first step of this project is to import the data from the csv.file provided by the Gowalla Website. Users can select the dataset file from the disk and load it into the system by storing different data in different vectors in a class which is specially designed to transmit its address in the whole project.

The loading of the data is implemented via multi-thread in order to prevent the program from being stuck. The starting signal is emitted to sub-thread, and it runs the working class. A QProgressBar is also added to show the loading progress.

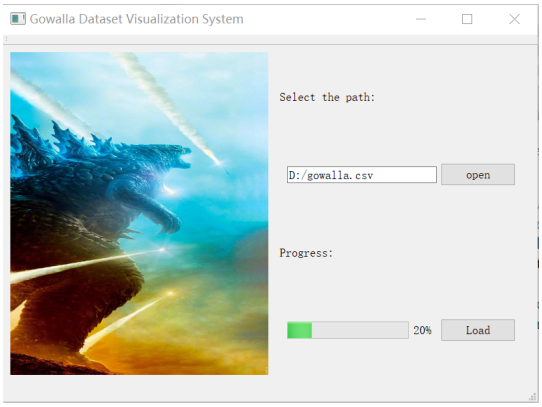


Figure 1: Import the Data

2.2 Basic Data Visualization

After the import of the data, a new window which will show us all the following functions will appear.

The first function is to realize the display of the spatio-temporal information. In this part, I wanted to draw a line that can depict the number of checking-ins as well as the the changing tendency.

To make the visualization more flexible and user-friendly, I added some widgets to tune the parameters for users to change the range conveniently. QSpinbox and QDateTimeEdit are used to make it available by connecting the ui interface to a specific struct stored in the main class.

Figure 2: Visualization of the number-time line

To draw a meaningful line, I added a QCombobox for users to choose the precision of the statistical result. Allowing for the fact that our time is stored in the form as seconds, I set 3 values as 604800, 86400, 3600, representing week, day and hour. Starting from the time we get from the ui interface, I count the number of checking-ins due to the given conditions and add simultaneously in the counting array. QSplineSeries is used to smooth the line we get through interpolation, and the picture can be displayed via QChartView. To make the x-axis more clearly, I substituted the QValueAxis with a QDateTimeAxis.

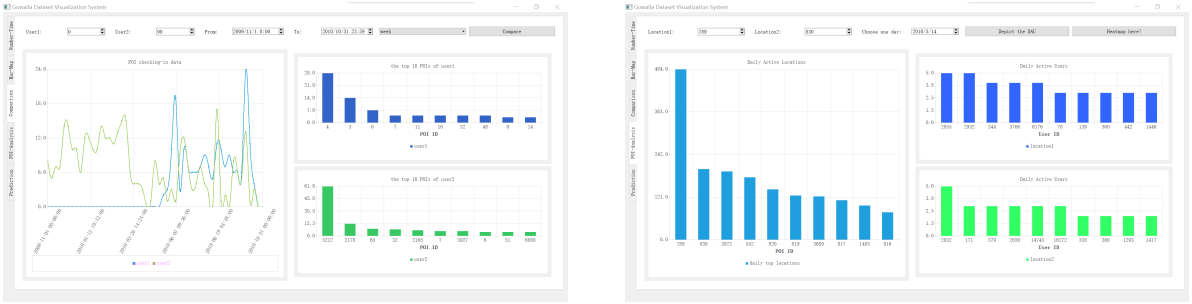
In this part, you can also select a specific POI or user-ID, and widgets for latitude and longtitude are also provided for you to select a specific GPS range. The filtering is implemented via multi-thread.

2.3 Comparison

Besides displaying the basic temporal and spatial line, I also implemented some comparison functions for users and POIs respectively.

The module can be divided into two parts. The first one is to compare the checking-in information of two users. Similar to the function in the 2.1 part, I added the function of plotting two number-time lines in one chart, using different colors. Parameters can also be tuned in your interest.

Moreover, I also depicted the top 10 popular POIs visited by the selected two users in the form of bar charts via QBarSeries. You can make some comparisons and analysis based on this.



(a) users (b) POIs

Figure 3: Comparisons

The second comparison is centered around two specific POIs on a given day. You can see the daily popular POIs and corresponding daily active users of a POI.

2.4 Prediction

The prediction module is divided into two parts, one is to predict a given POI's expected checking-in number on Oct 18, 2010, Oct 19, 2010, and Oct 20,2020. According to some information on the Gowalla Website and some analysis on my own based on 2.1 part, I take the data before Oct 18 as valid statistics, and choose a time span long as about 110 days for computation.

For prediction, I adopted the simple exponential smoothing method. In fact, two estimation method are often adopted.The first one is to let the latest value equal the last observation value regardless of others; the other one is to take the average as the value. Considering the time effect, we choose simple exponential smoothing, which has both features to make predictions. In this project, we choose the average value of the first three days as an initial value, and then take one day as a step, calculating the average value as a new initial value until the time moves to Oct 18, 2010.

Considering periodicity, based on previous analysis that one week is a cycle. We store data in three different vectors at a 7-day interval, and predict respectively.

For the prediction of a user's next point of interest. I stored a given user's visited POI IDs in time order, and adopt a Markov Chain to analyze the locus on the basis of probability.

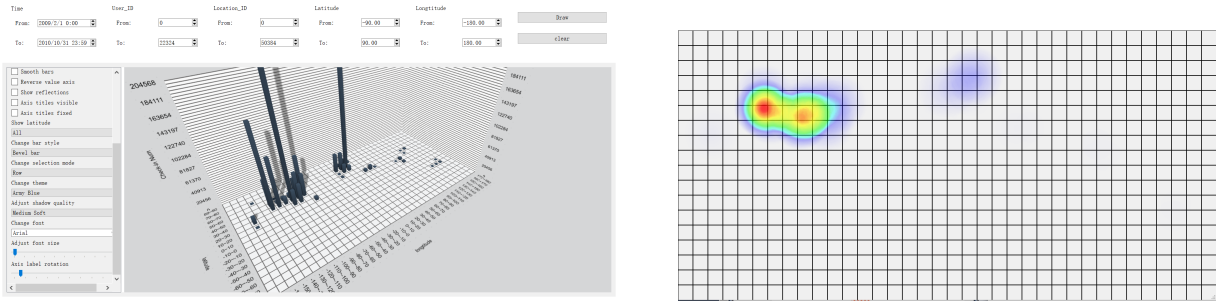
Suppose one user visited n different POIs, we define a n*n matrix for probability calculation. If one POI to another POI is recorded in the chain, then the corresponding value in the matrix is set to 1. After a traverse, we make normalization in every row. After a large number of iterations, the matrix can attain a stable values, that is the probability we want. Then the user's expected next POI can be found.

2.5 Distribution Map

In this module, I wanted to show the data distribution on a map. But as time is limited, I failed to import real world map into this project via QML. I chose to realize it in a simpler way. I implemented a 3D bar chart to show the distribution of Gowalla checking-in numbers. The

parameters can also be tuned as you like. The latitude and longtitude are mapped on the flat. I took 10 degrees as a statistical region, the height of each representing the number.

Also, I implemented a heatmap, which can depict the popularity of a range of locations via different colors. This function is placed in the comparison module. When analyzing daily popular POIs, you can have a look at the heatmap. The size of the map window is fixed, and every cell's length represents 10 degrees.



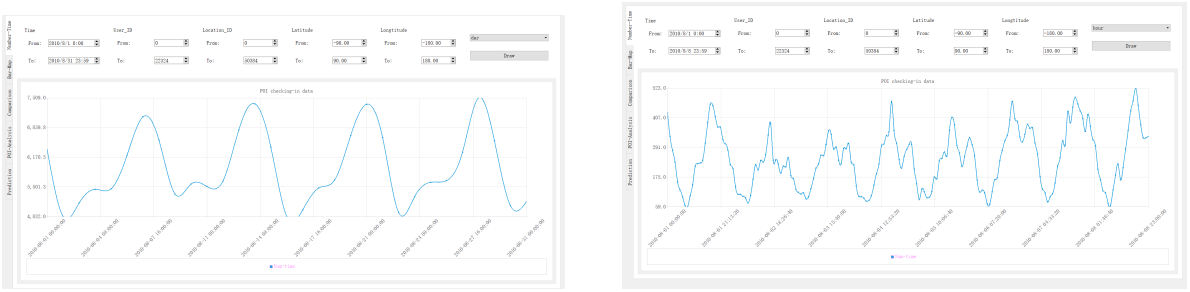
(a) bar (b) heat

Figure 4: map visualization

3 Results

3.1 Spatio-temporal Distribution Patterns

From the visualization result, we can discover that most of the time, the checking-in number was changing in a specific cycle in general, with clear peaks and values. For example, the cycle of week shows peaks mostly on Saturday while on workdays the checking-in number is relatively lower. Moreover, the number reached its peak mostly in the afternoon and at evening. So we could draw a conclusion that people tended to check in their gowalla accounts at weekends and in the afternoon.



(a) week observation (b) day observation

Figure 5: spatio-temporal visualization

And from the overall perspective, the checking-in number started its rapid increase from September, 2009 and reached its highest in March, 2010 and then keep a stable situation until the website closed in the end of October, 2010.

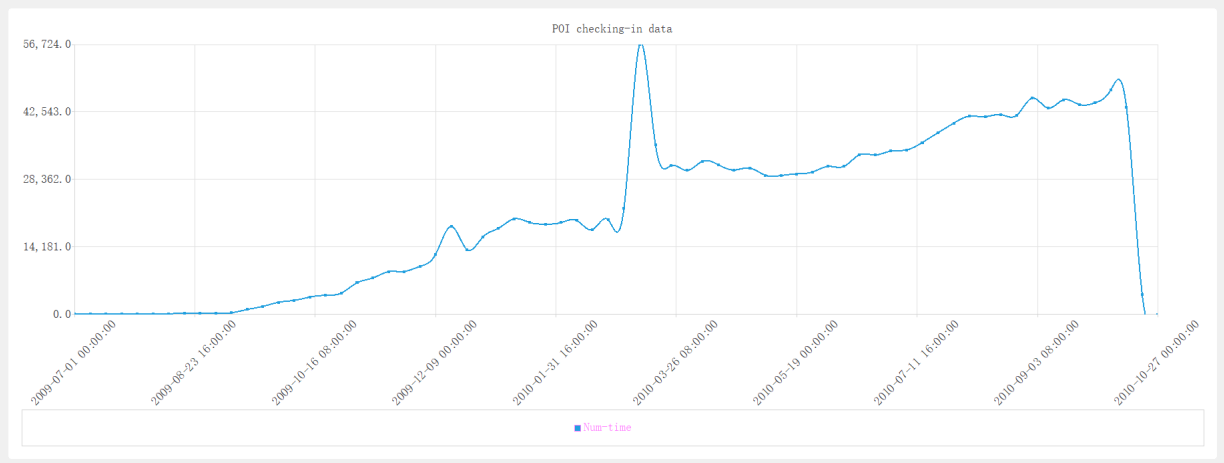


Figure 6:

3.2 Location Preferences

According to the map analysis and some results of comparisons, we find that the popular POIs are mostly distributed in latitudes between 30 and 40, longtitudes between -100 and -90 and latitudes between 50 and 60, longtitudes between 10 and 20.

4 Discussion

4.1 Performance of Application

Many functions of this project are implemented based on multi-thread. Without it, the dataloading would take more than one minute with no responses. With multi-thread, the interface works more smooth and only take 20 seconds to import the data. And during development, I discover that the display of progressbar takes far more time than other computations.

4.2 Expand Ideas

1. Dynamic real maps can be imported into the map combined with heatmap through QML. We can depict user's locus on the map and carry on deeper analysis.
2. The website can rank users referring to a standard to stimulate users to contribute to and interact with the website community. For example, the checking in numbers and their covering ranges and some other relevant elements.

5 Acknowledgement

I would like to express my sincere gratitude to everyone who has helped me during my development.

Thanks to Dr.Jin and Dr.Zhao for their well-prepared teaching and tender guidance during the whole semester.

Thanks to TAs for their hard work to keep our class progressing smoothly.

Thanks to my friend who discuss with me and answer my questions for their patience and kindness.