Location recommendation of venues based on data from Foursquare

Zijie Huang

December 28, 2019

1. Introduction

1.1. Background

Nowadays, as the actual users of social networks, people's living needs are constantly growing, and there are numerous social networking websites and software. Location-Based Social Networks (LBSN), as one of the characteristics of social networks, is also accompanied by the development of these software. Compared with others ordinary social networks, the LBSN is more prominent in that it can reflect the exact information of users in time and space, so as to obtain the specific real-time or certain-time location information of these users, then using these to understand the user's behaviour and preferences.

1.2. Problem

However, with the explosive development of data now, information overloading will also cause some unnecessary trouble. Different users have different interests and hobbies, and even the same users will behave differently in different events. Therefore, it is important for social networks to make personalized recommendations about what they are about to do. Personalized location recommendation aims to mine the user's check-in information through machine learning, to obtain the personalized features of interest of each user, and then combine some current information to predict and recommend. The main purpose is not only to make the user more convenient, but also to produce a series of commercial value.

1.3. Interest

This report mainly focuses on when the user does not know where to make a new venue and which kind of venue is the best choice, he needs to make specialised recommendations based on the previous queries.

2. Data acquisition and cleaning

2.1. Source

This article uses the POI data set of the user's check-in records in Toronto on Foursquare. There are 2,223 POI useful data out of 3,000, and each record contains the location category, latitude and longitude, and venue.

2.2. Data cleaning

I have already scraped the postal code of Toronto from Wikipedia and several POIs from Foursquare. In order to improve the efficiency, I have removed some useless data, and then get 2,223 queries and 270 categories.

```
In [58]: tor_venues
```

2210	Kingsview Village, Martin Grove Gardens, Richv...	43.688905	-79.554724	Ital Pizza	43.690136	-79.559979	Pizza Place
2211	Kingsview Village, Martin Grove Gardens, Richv...	43.688905	-79.554724	Rogers	43.692620	-79.557402	Mobile Phone Shop
2212	Albion Gardens, Beaumond Heights, Humbergate, ...	43.739416	-79.588437	Subway	43.742421	-79.589471	Sandwich Place
2213	Albion Gardens, Beaumond Heights, Humbergate, ...	43.739416	-79.588437	Shoppers Drug Mart	43.741685	-79.584487	Pharmacy
2214	Albion Gardens, Beaumond Heights, Humbergate, ...	43.739416	-79.588437	Popeyes Louisiana Kitchen	43.741202	-79.584545	Fried Chicken Joint
2215	Albion Gardens, Beaumond Heights, Humbergate, ...	43.739416	-79.588437	Sheriff's No Frills	43.741696	-79.584379	Grocery Store
2216	Albion Gardens, Beaumond Heights, Humbergate, ...	43.739416	-79.588437	The Beer Store	43.741694	-79.584373	Beer Store
2217	Albion Gardens, Beaumond Heights, Humbergate, ...	43.739416	-79.588437	Pizza Pizza	43.741569	-79.584489	Pizza Place
2218	Albion Gardens, Beaumond Heights, Humbergate, ...	43.739416	-79.588437	McDonald's	43.741757	-79.584230	Fast Food Restaurant
2219	Albion Gardens, Beaumond Heights, Humbergate, ...	43.739416	-79.588437	Sunny Foodmart	43.741840	-79.590561	Grocery Store
2220	Albion Gardens, Beaumond Heights, Humbergate, ...	43.739416	-79.588437	LCBO	43.741508	-79.584501	Liquor Store
2221	Northwest	43.706748	-79.594054	Economy Rent A Car	43.708471	-79.589943	Rental Car Location
2222	Northwest	43.706748	-79.594054	Saand Rexdale	43.705072	-79.598725	Drugstore

```
print('There are {} uniques categories.'.format(len(tor_venues['Venue Category'].unique())))
```

```
There are 270 uniques categories.
```

2.3. Feature selection

As a common method for text feature extraction, one-hot uses N-bit status register to encode N states, each state has independent register bits, and only one of these register bits is valid. It can help to solve the problem that the classifier has difficulty processing discrete data.

After this step, the data frame we got is this:

61]:

	Yoga Studio	Accessories Store	Afghan Restaurant	Airport	Airport Food Court	Airport Gate	Airport Lounge	Airport Service	Airport Terminal	American Restaurant	Antique Shop	Aquarium	Art Gallery	Arts & Crafts Store	Asian Restaurant	Athletics & Sports	Auto Garage	Auto Workshop	BBQ Joint	Baby Store	Bagel Shop	Bakery	Ban
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
4	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	

```
: tor_onehot.shape
```

```
62]: (2223, 270)
```

3. Exploratory data analysis

3.1. Generate venue frequencies for each area

Based on the one-hot table above, we can contribute the frequency, or the weight of these venues in the area. The higher it gets, the more common this venue be in this area.

```
----Adelaide, King, Richmond----
              venue  freq
0       Coffee Shop  0.08
1              Café  0.05
2        Steakhouse  0.04
3               Bar  0.04
4   Thai Restaurant  0.03
5  Asian Restaurant  0.03
6      Burger Joint  0.03
7        Restaurant  0.03
8  Sushi Restaurant  0.03
9       Salad Place  0.03


----Agincourt----
                       venue  freq
0  Latin American Restaurant   0.2
1                     Lounge   0.2
2             Breakfast Spot   0.2
3             Clothing Store   0.2
4               Skating Rink   0.2
5               Yoga Studio    0.0
6   Mediterranean Restaurant   0.0
7         Miscellaneous Shop   0.0
8   Middle Eastern Restaurant  0.0
9         Mexican Restaurant   0.0
```

3.2. Recommendation making

After that, the frequency shows the common places in each area, what we can do now is to simply combine it with the area together to make a new table. Here I have chosen the top 5 common venue for the areas.

	Neighborhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue
0	Adelaide, King, Richmond	Coffee Shop	Café	Bar	Steakhouse	Salad Place
1	Agincourt	Latin American Restaurant	Skating Rink	Clothing Store	Breakfast Spot	Lounge
2	Agincourt North, L'Amoreaux East, Milliken, St...	Park	Playground	Discount Store	Dance Studio	Deli / Bodega
3	Albion Gardens, Beaumond Heights, Humbergate, ...	Grocery Store	Pizza Place	Fried Chicken Joint	Sandwich Place	Liquor Store
4	Alderwood, Long Branch	Pizza Place	Sandwich Place	Coffee Shop	Skating Rink	Gym
5	Bathurst Manor, Downsview North, Wilson Heights	Coffee Shop	Shopping Mall	Pharmacy	Supermarket	Sushi Restaurant
6	Bayview Village	Café	Chinese Restaurant	Japanese Restaurant	Bank	Dessert Shop
7	Bedford Park, Lawrence Manor East	Italian Restaurant	Coffee Shop	Fast Food Restaurant	Sushi Restaurant	Comfort Food Restaurant
8	Berczy Park	Coffee Shop	Cocktail Bar	Seafood Restaurant	Farmers Market	Beer Bar
9	Birch Cliff, Cliffside West	Café	College Stadium	Skating Rink	General Entertainment	Women's Store
10	Bloordale Gardens, Eringate, Markland Wood, Ol...	Park	Coffee Shop	Café	Pet Store	Pharmacy
11	Brockton, Exhibition Place, Parkdale Village	Café	Coffee Shop	Breakfast Spot	Pet Store	Stadium

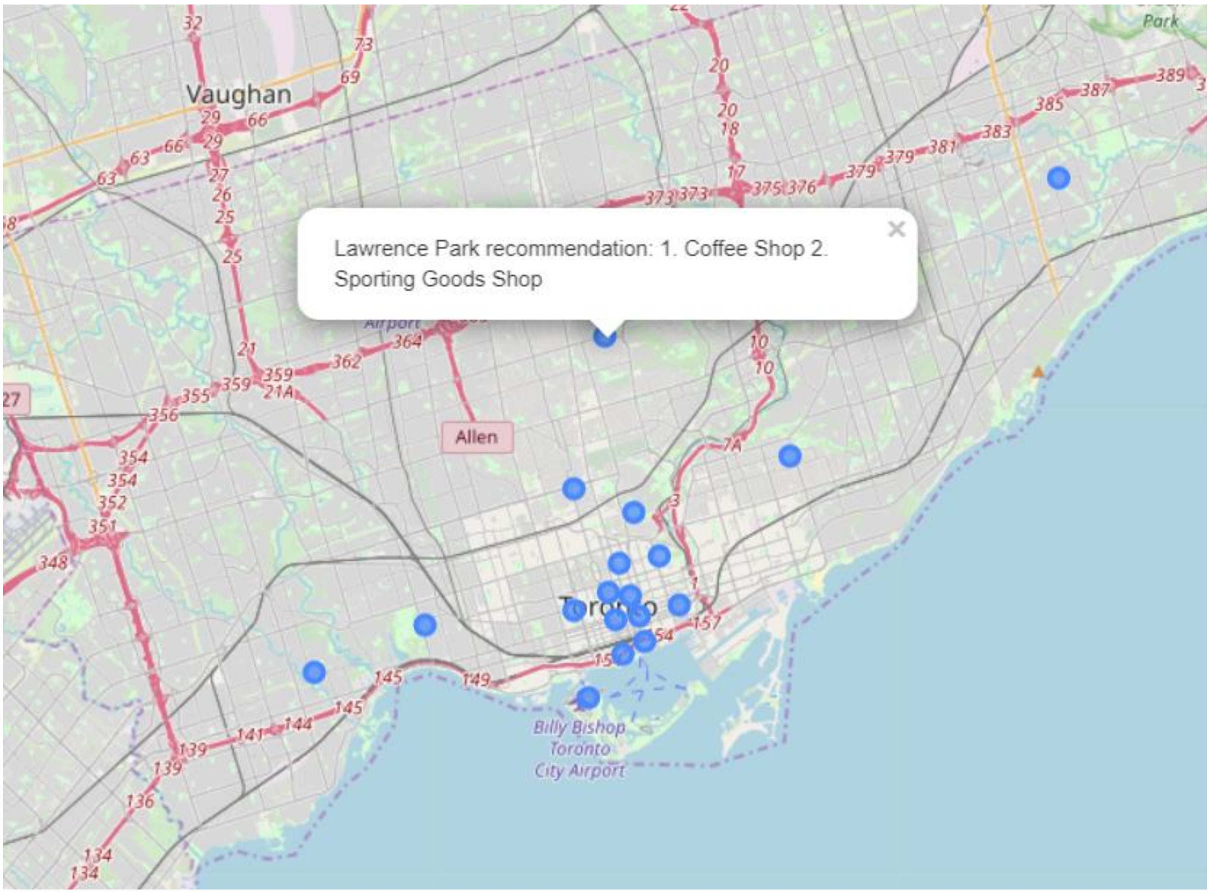
Also, don't forget to add the latitude and longitude from the area if you want to build a map to show this, here is another example.

	PostalCode	Borough	Neighborhood	Latitude	Longitude	RecommendedOne	RecommendedTwo
0	M1B	Scarborough	Rouge, Malvern	43.806686	-79.194353	Fast Food Restaurant	Women's Store
1	M1C	Scarborough	Highland Creek, Rouge Hill, Port Union	43.784535	-79.160497	Bar	Women's Store
2	M1E	Scarborough	Guildwood, Morningside, West Hill	43.763573	-79.188711	Rental Car Location	Pizza Place
3	M1G	Scarborough	Woburn	43.770992	-79.216917	Coffee Shop	Korean Restaurant
4	M1H	Scarborough	Cedarbrae	43.773136	-79.239476	Fried Chicken Joint	Thai Restaurant
5	M1J	Scarborough	Scarborough Village	43.744734	-79.239476	Spa	Playground
6	M1K	Scarborough	East Birchmount Park, Ionview, Kennedy Park	43.727929	-79.262029	Discount Store	Chinese Restaurant
7	M1L	Scarborough	Clairlea, Golden Mile, Oakridge	43.711112	-79.284577	Bakery	Bus Line
8	M1M	Scarborough	Cliffcrest, Cliffside, Scarborough Village West	43.716316	-79.239476	Motel	American Restaurant
9	M1N	Scarborough	Birch Cliff, Cliffside West	43.692657	-79.264848	Café	College Stadium
10	M1P	Scarborough	Dorset Park, Scarborough Town Centre, Wexford ...	43.757410	-79.273304	Indian Restaurant	Pet Store
11	M1R	Scarborough	Maryvale, Wexford	43.750072	-79.295849	Middle Eastern Restaurant	Auto Garage

4. Results

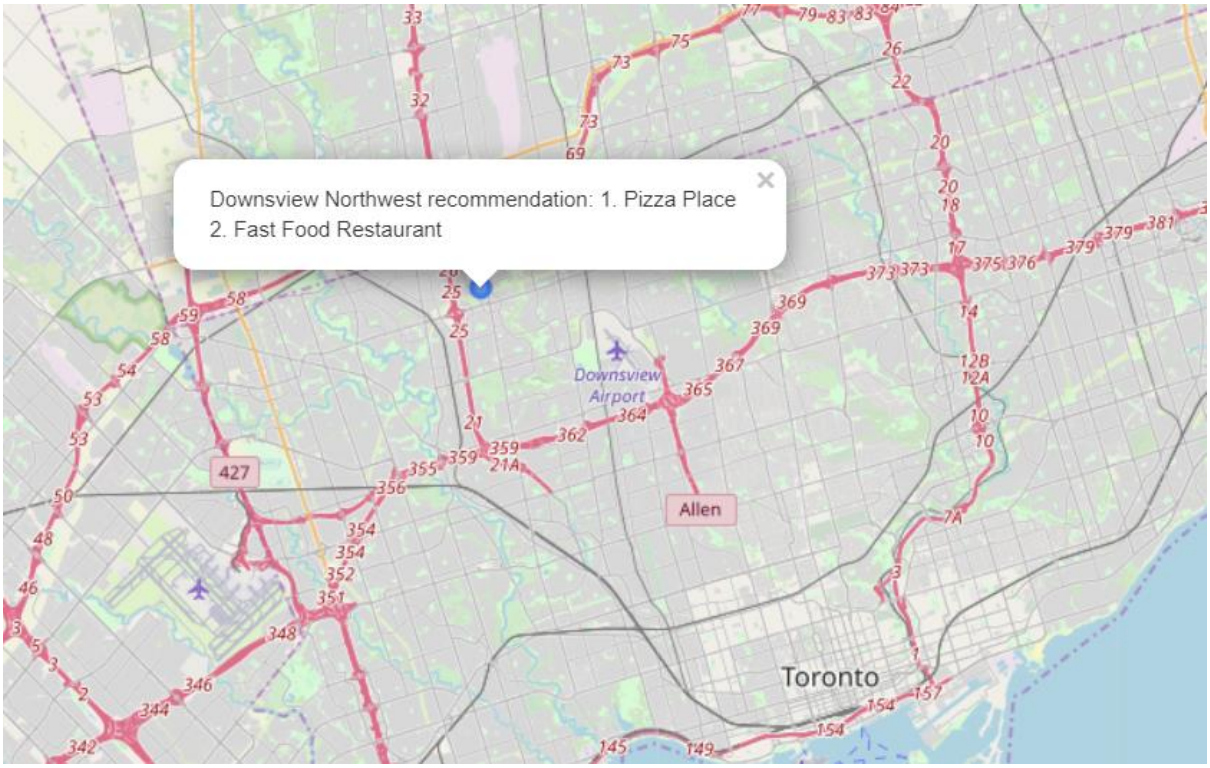
4.1. Get position recommendation

If we want to build a new coffee shop, where is a good place? Just put the venue into this model, and it shows the recommended places in the map.



4.2. Get venue recommendation

This model also can recommend what is the most popular venue in a given area. If we want to know which venue is the best recommended in the given postal code “M3N”, or name “Downsview Northwest”, we can find that:



5. Discussion

In this paper, based on the check-in data based position recommendation model adopted for the current problems, there are still many improvements in calculation details and data processing that can be improved. Future work can be further studied in the following directions:

1. The framework constructed in this paper is general and can be further expanded in many aspects according to the needs of actual data. For example, we can study the influence of social groups on location preferences by collecting social relationships between various users. This will make the model more realistic, and more accurately mine the user's personal preferences and daily habits.
2. The calculation process of the optimization algorithm. We can further improve the steps and logic of the algorithm, further optimize the algorithm, reduce the program running time, reduce system loss, and improve algorithm performance. For example

we can standardise the numeric data and label the text to get a higher processing speed.

3. In addition, due to the limited data in the database used in this article, the position information only has latitude and longitude, but only two-dimensional positions. The calculated results cannot be directly applied to three-dimensional buildings such as department stores; There is a certain lack of information and some data cannot be used. In the next research, we will conduct further research on the selection and preprocessing of data, so as to dig into richer content, make more satisfactory recommendations for customers.