Recommended restaurant in N.Y

Aug 5, 2020

1. Introduction

My client who operates some restaurants in New York often mailed to customers who ever visited there in order to drive sales. But his sales volume has not increased more than he expected.

Now he thinks that any strategy is needed to invite new customer. So, I propose finding prospective new customer on Forsqare users and advertise them by any means.

2. Data

Forsquare provides venues and users who liked or listed it, so that we can assume what kind of people is interested in what kind of venue in a certain extent. These data can be utilized to recommend venues which he has not yet visited.

(Unfortunately, venues which has been visited by users actually are not requested by other than themselves.)

2.1. Categories of restaurants

On Forsqure, there are many kind of restaurant categorized, American, Japanese, Chinese,etc. To get venues categorized restaurants afterword, use API and get categories list.

API response is dictionary format, and category records (id, name) are hierarchized.

According to the spec of category data(https://developer.foursquare.com/docs/build-with-foursquare/categories/) , 3 categories are enough for food category.

	category1_id	category1_name	category2_id	category2_name	category3_id	category3_name
0	503288ae91d4c4b30a586d67	Afghan Restaurant	NaN	NaN	NaN	NaN
1	4bf58dd8d48988d1c8941735	African Restaurant	4bf58dd8d48988d10a941735	Ethiopian Restaurant	NaN	NaN
2	4bf58dd8d48988d14e941735	American Restaurant	4bf58dd8d48988d157941735	New American Restaurant	NaN	NaN
3	4bf58dd8d48988d142941735	Asian Restaurant	56aa371be4b08b9a8d573568	Burmese Restaurant	NaN	NaN
4	4bf58dd8d48988d142941735	Asian Restaurant	52e81612bcbc57f1066b7a03	Cambodian Restaurant	NaN	NaN

<df_cat>

2.2. Restaurants in NY

Get venue API is used. Responded venue' property has only category class, so merge previously created df_cat.

2.3. Tips and Lists

Get venue's detail API is used. Assumption is that tips or lists is created because he liked there. So users who posted tips or created lists are corrected in venue's detail and df_tips and df_lists can be created.

Get Venues Details API is premium call (500 calls per day) so executed daily batch on IBM Watson Studio.

Venues on Tips and Lists will be scored 5, and 4.

2.4. Liked

Using users, which extracted from tips and lists, Venues with flag 'liked' them can be called.

The number of calls are very large. Because API call is limited 500 calls per hour, so so executed hourly batch on IBM Watson Studio.

Venues on Tips and Lists will be scored 3.

2.5. Venues with same category as his existing favored venues

Of all venues, venues with same category or same super-category as ones on his tips, lists, and liked will be also populated.

Venues on same category and same super-category will be scored 2 and 1.

Now these data are merged and scored data will be created.

(15218507, 3)

	user_id	venue_id	score
0	1265687	582a1b095f78034bcfb69186	5
1	192962359	582a1b095f78034bcfb69186	5
2	2868030	57b8f763cd10f71921e464b5	5
3	59130915	51ec708d498e58679aa83bca	5
5	35551887	541f4507498e3a898af35a90	5

3. Methodology

Over 15 million scoring data with for about 8 thousand users and 4 thousand venues(food category) has been collected, so I think we can predict scores of unevaluated venues by collaborative filtering.

- liked venues as 3
- venues on lists as 4 and ones on tips as 5
- venues same category as on lists and tips as 2
- venues same super-category as on lists and tips as 1

	user_id	venue_id
score		
1	1621397	1621397
2	13576788	13576788
3	4777	4777
4	10060	10060
5	5485	5485

< Counts for each score >

3.1. Collaborative filtering

Predict users' ratings for venues, using machine learning algorism (Collaborative filtering).

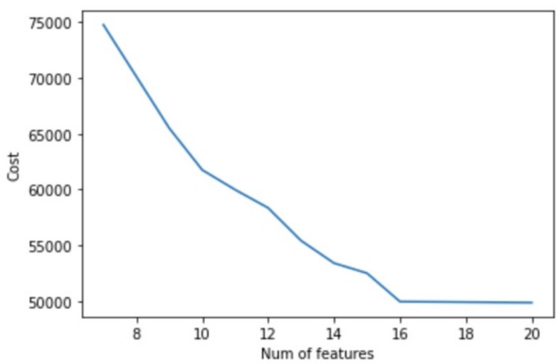
3.1.1. Functions

Use the algorism which updates the rate of venues' features and the rate of users' features during training.

Data will be normalized by standard deviation.

3.1.2. Training

Cost for each combination of the number of feature and the amount of bias was as followings. So number of features will be set **20.**



3.1.3. Predict

Predicted score matrix for was created as bellow.

	32	33	37	56	59	61	135	140	144	209	...	580677
5ba682d616ef670039f6d605	2.014082	1.987447	2.024548	2.009573	1.999409	1.989089	2.010838	2.034907	2.026080	1.990084	...	2.033
5c13bf838afbe0002de55061	1.999777	2.009438	2.003895	2.006324	1.996612	2.007620	2.010113	2.024210	2.015391	2.005031	...	2.008
4cb218c7ef1b37041ca24800	2.005399	1.997427	2.002186	2.008171	2.001285	1.998127	2.001735	2.011519	2.026425	2.008373	...	2.023
57fa6c67498e23e3cf1861f4	1.991532	1.997121	1.999090	2.003478	1.999316	1.997626	2.005188	2.018935	2.029585	2.011068	...	2.017
5615a472498ed90cc0623b3c	2.002414	2.006484	2.006006	2.005894	2.005643	2.005005	2.002959	2.022838	2.021965	2.003634	...	2.018
...	...	...	...	...	...	...	...	...	...	...	...	...
5adcbd038173cb433ff5350d	1.380004	0.906656	1.163593	1.013677	1.153813	1.260904	0.658151	0.948888	0.349389	1.262704	...	1.294
52f66aa6498e08f2b069ab80	1.010813	1.001961	0.999621	1.014716	1.026965	1.003386	0.999194	1.000771	0.972044	1.015611	...	1.005
596fb23c31ac6c49a7f6b6c4	1.001377	0.998714	1.005378	1.002443	1.010739	1.001027	1.003945	1.005788	1.000046	1.001039	...	1.012
4cddd9dadb125481c4f22cce	1.115300	0.622533	1.091012	0.939870	1.112741	1.031712	0.704897	0.963842	0.837573	1.217481	...	1.323
590e07a81bc704026f2e9400	1.142625	0.713616	1.119190	0.901105	0.976738	1.053903	0.756639	0.857028	0.765521	1.084026	...	1.264

3428 rows × 7780 columns

4. Result

4.1. Compare score and prediction

To make sure if already high scored restaurants are predicted high as well, one user' top 5 scored venue and corresponding prediction.

	user_id	venue_id	score	venue_name	category_id	category_name
0	1265687	582a1b095f78034bcfb69186	5	Sami's Kebab House	503288ae91d4c4b30a586d67	Afghan Restaurant
1	1265687	56e064dc498e1ba17ec18b6f	5	Fu Yuan	56aa371be4b08b9a8d57350b	Food Stand
2	1265687	5d7d2f62774999000703bdcb	3	Yin Ji Chang Fen 銀記腸粉店	52af3a7c3cf9994f4e043bed	Cantonese Restaurant
3	1265687	5a13402d286fda3f038703e0	3	Le Privé	4bf58dd8d48988d10c941735	French Restaurant
4	1265687	58225a8f001d8755c9c4bcd1	3	Sweetcatch Poke Bar	5bae9231bedf3950379f89d4	Poke Place

<Top 5 scores of user id '1265687' ordered by score>

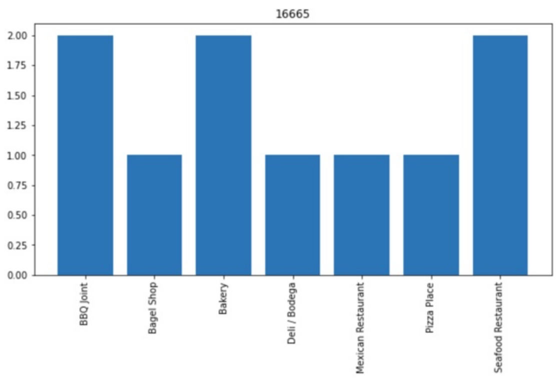
	venue_id	venue_name	category_id	category_name
2	582a1b095f78034bcfb69186	Sami's Kebab House	503288ae91d4c4b30a586d67	Afghan Restaurant
218	5a13402d286fda3f038703e0	Le Privé	4bf58dd8d48988d10c941735	French Restaurant
904	56e064dc498e1ba17ec18b6f	Fu Yuan	56aa371be4b08b9a8d57350b	Food Stand
2993	58225a8f001d8755c9c4bcd1	Sweetcatch Poke Bar	5bae9231bedf3950379f89d4	Poke Place
3558	5d7d2f62774999000703bdcb	Yin Ji Chang Fen 銀記腸粉店	52af3a7c3cf9994f4e043bed	Cantonese Restaurant

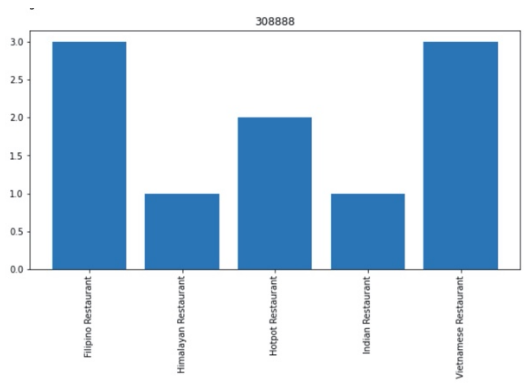
<Top 5 predictions of user id '1265687' ordered by score>

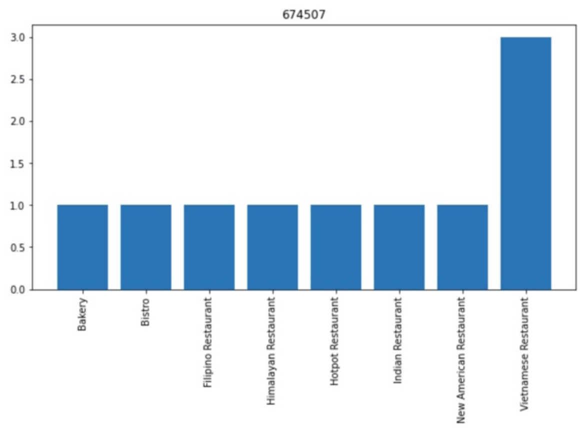
5 scored restaurants were highly ranked on prediction as well.

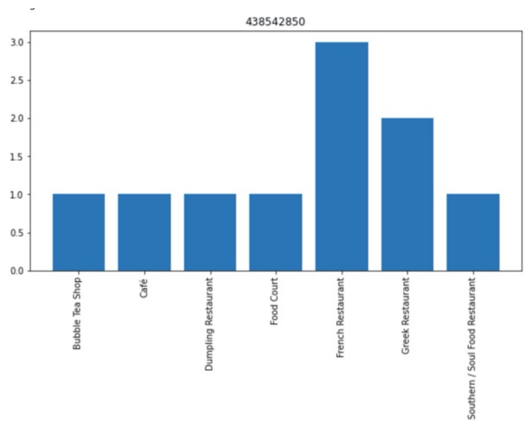
5. Discussion

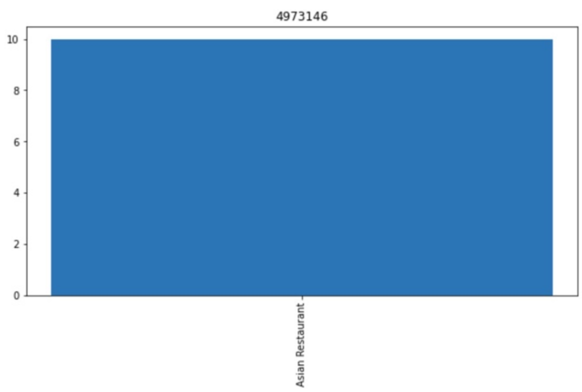
5.1. 10 users were randomly picked up and their top 30 venues were ordered by their scores to look up how category of their recommended venues are varied.

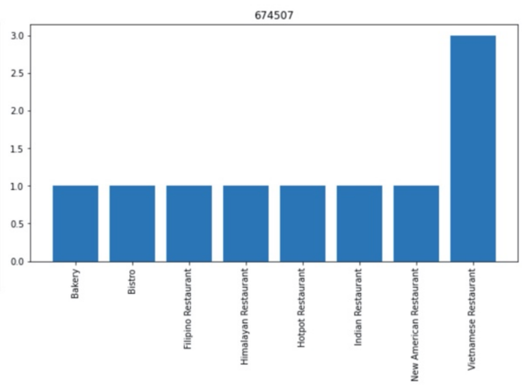










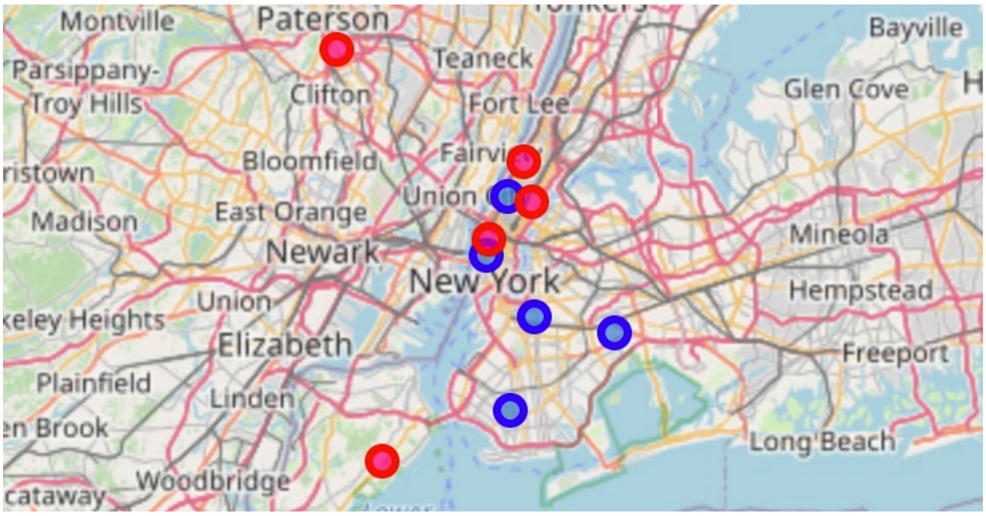


It is apparent that their recommended restaurants belong to same category, but in some cases they cover a variety of categories.

5.2. Identify location on map

Here, we focus on user '53052590', whose cover a variety of categories recommended, and identify them on map

Blue popups are venues already scored by him, and Red popups are ones predicted.



6. Conclusion

We could get recommendation somehow. We hope we could propose users to visit their recommendation and they got back to us! For more precise prediction, we need more data as followings

1. **Much more cross evalutions. (Scoring was done base on "tips", "listed", 'liked' flag and category, however, there were not enough cross evaluations for collaborative filtering to reach success.)**

2. **Venues they visited. (As for now, API does not support it)**