Point-of-interest recommendation for travelers. Special Edition of the Text REtrieval Conference (TREC) Contextual Suggestion Track at RuSSIR 2015.

'MAD IT' team: Maria, Andrey, Dmitry, Ivan, Tatiana

August 27, 2015

Overview

The goal of the hackathon was to elaborate recommendation for three users based on the their history of visited places. Each history record consists of tags user marked the place and rating (from -5 to 5). The complete description of the task and dataset could be found here: http://plg.uwaterloo.ca/~claclark/russir2015.

Our approach is based on estimating relatieve weight of each tag user assigned to visited place and using this weigths to calculate rating for new place by averaging values associated with each tag.

Data munging

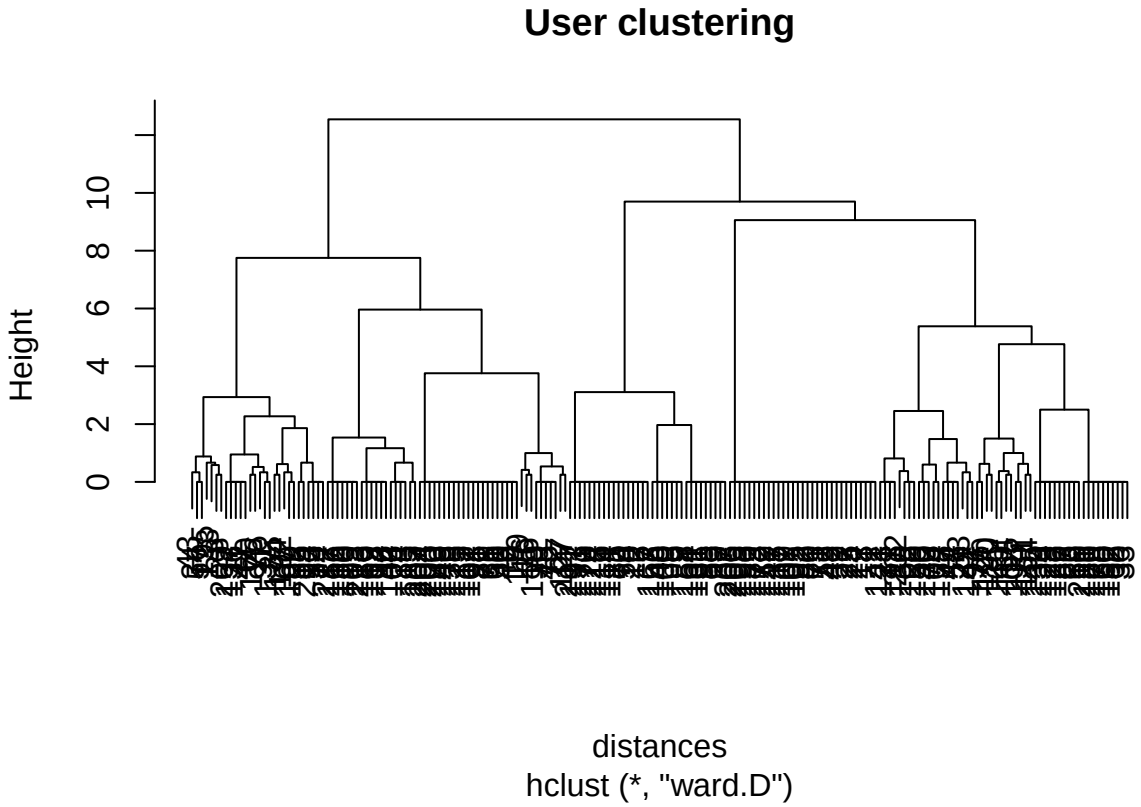
One of the most serious problem faiced our team was issues with test data. The dataset emulates most of the real world problems such as:

- original data provided in JSON format and need to be converted before analysis,
- missed data (there are 214 profiles but only 194 of them are complete; one of the target user doesn't have any ratings for visited place; absence of tags and ratings for Saint-Petersburg places),
- inconsistent data (e.g. gender might be "mail", "femail" or "f"),
- data duplication (e.g. there are many very similar categories: "Art Galleries", "Museums" and "Fine Art Museums" or "Coffee" and "Cafes" etc.),
- small number of users datasets,
- dataset of ratings is very sparse.

Some of these issues were solved by using R, others were fixed manually, but the remained problems couldn't be solved due to the lack of information.

Split users into cohorts

Users were split into five clusters by theirs gender, group, trip season, duration and type. Although it wasn't obvious while looking at hierarchical clustering dendogram, we choose five clusters. The goal was to produce ad hoc solution such that target users with similar traits end up in the same cluster, but do not assigh the same cluster to the all three target users:



Target users:

	id	age	gender	group	season	duration	trip.type	cluster
212	1234567	50	male	friends	summer	longer	holiday	3
213	1234568	47	male	friends	autumn	longer	holiday	3
214	1234569	30	female	alone	summer	weekend trip	holiday	2

Calculate target users preferences:

Weights of tags for user with *id=1234567* were calculated by averaging ranks of tags for *cluster 3* because there are no any tag for this user:

	newtags	rating
60	nature	3.159575
65	tropical	3.000000
49	cafs	2.771930
62	restaurant	2.628141
52	drinking	2.586956
45	art	2.581967
58	museum	2.549618

	newtags	rating
55	history	2.500000
61	relax	2.462428
46	authentic	2.336956
57	leisure	2.242991
44	active	2.177632
59	music	2.160000
50	city	2.104651
64	sport	2.026316
63	shopping	2.015544
51	cold weather	2.000000
56	hot weather	2.000000
66	warm weather	2.000000
53	family friendly	1.980392
54	fast food	1.900000
48	business	1.684211
67	wellness	1.642857
47	budget friendly	1.000000

Weights of tags for user with *id=1234568*:

	newtags	rating
2093	art	3.000000
2094	authentic	3.000000
2099	museum	2.750000
2098	leisure	2.666667
2103	restaurant	2.451613
2092	active	2.333333
2096	drinking	2.000000
2101	nature	2.000000
2105	sport	2.000000
2100	music	1.800000
2102	relax	1.800000
2095	city	1.500000
2104	shopping	1.333333
2097	fast food	0.000000

Weights of tags for user with *id=1234569*:

	newtags	rating
2107	art	4.000000
2109	leisure	4.000000
2111	music	3.000000
2112	relax	3.000000
2113	restaurant	2.916667
2110	museum	2.500000
2114	sport	2.000000
2106	active	0.000000
2108	authentic	0.000000

Make predictions

Recommendations for user with *id=1234567*:

	name	url
102	Le Tour de Vin	https://foursquare.com/v/le-tour-de-vin/5370e6d8498e666a1bfe1c0
24	Peterhof	http://www.tripadvisor.com/Attraction_Review-g1207894-d261586
15	Park of the 300th anniversary of St. Petersburg	https://foursquare.com/v/%D0%BF%D0%B0%D1%80%D0%BA-30
28	Cafe Abajour	http://www.tripadvisor.com/Restaurant_Review-g298507-d318393
36	Bushe	http://www.tripadvisor.com/Restaurant_Review-g811323-d655745
38	Bize	http://www.tripadvisor.com/Restaurant_Review-g298507-d376977
43	Zoom Cafe	http://www.tripadvisor.com/Restaurant_Review-g298507-d799299
45	Palkin	http://www.tripadvisor.com/Restaurant_Review-g298507-d110599
58	Dreamers	https://foursquare.com/v/%D0%BC%D0%B5%D1%87%D1%82%D
61	Manneken Pis	http://www.tripadvisor.com/Restaurant_Review-g298507-d611929

Recommendations for user with *id=1234568*:

	name	url
16	Grand Maket Russia Interactive Museum	http://www.tripadvisor.com/Attraction_Review-g298507-d3396387-R
26	Faberge Museum	http://www.tripadvisor.com/Attraction_Review-g298507-d6580510-R
44	Mikhailovsky Theatre / Mikhailovsky Theatre	https://foursquare.com/v/%D0%BC%D0%B8%D1%85%D0%B0%D0
68	Grand Model Russia	https://foursquare.com/v/%D0%B3%D1%80%D0%B0%D0%BD%D0
32	Percorso	http://www.tripadvisor.com/Restaurant_Review-g298507-d5247712-R
41	Severyanin	http://www.tripadvisor.com/Restaurant_Review-g298507-d2645717-R

	name	url
59	Russian Empire	http://www.tripadvisor.com/Restaurant_Review-g298507-d3764629-R
62	Caffe Italia	https://foursquare.com/v/caffe-italia/4bba3d7153649c7498cc48fb
77	Namaste	http://www.tripadvisor.com/Restaurant_Review-g298507-d7056042-R
94	Saint Scalpelburg	https://foursquare.com/v/saint-scalpelburg/4f0c6a63e4b02a8da65680

Recommendations for user with *id=1234569*:

	name	url
44	Mikhailovsky Theatre / Mikhailovsky Theatre	https://foursquare.com/v/%D0%BC%D0%B8%D1%85%D0%B0%D0
94	Saint Scalpelburg	https://foursquare.com/v/saint-scalpelburg/4f0c6a63e4b02a8da65680
20	State Academical Mariinskiy Theatre	http://www.tripadvisor.com/Attraction_Review-g298507-d7082919-R
33	Tar baths	https://foursquare.com/v/%D0%B4%D0%B5%D0%B3%D1%82%D1%
66	Venezia / Venice	https://foursquare.com/v/venezia–%D0%B2%D0%B5%D0%BD%D0%
22	Mikhailovsky Opera and Ballet Theater	http://www.tripadvisor.com/Attraction_Review-g298507-d2619507-R
84	Mariinsky Theatre / Mariinsky Theatre	https://foursquare.com/v/%D0%BC%D0%B0%D1%80%D0%B8%D0
21	GreenWood	https://foursquare.com/v/greenwood/55aaa65c498eeed921feb4da
16	Grand Maket Russia Interactive Museum	http://www.tripadvisor.com/Attraction_Review-g298507-d3396387-R
26	Faberge Museum	http://www.tripadvisor.com/Attraction_Review-g298507-d6580510-R

To make final submission in json format execute `source('response.R')`.

Conclusion

Although proposed method is pretty simple, it works quite good: we managed to guess at least one real preference of the user. The reason why it demonstrates such result might be the usage only relevant information. We use only ratings of similar users (averaging by cluster) or ratings of users themself. Obviously our solution lacks any reliable performance indicator: splitting original data into train and validation sets could help, but with such little dataset it most likely hurt performance.