CHICAGO BOOTH

The University of Chicago Booth School of Business

Digital and Algorithmic Marketing

BUSN 37304-85

Winter 2020

Project: Restaurant Recommendation for Groups

Team

Li Liu (12211053)

Suresh Govindaraj (10459129)

Xi Zhao

Hypothetical Readers: executives at Yelp who are in charge of developing new product features. They recently learned from a customer survey that Yelp was not easy to help customers decide when they dined out with people with different preferences. These customers were dissatisfied and stopped using the product. As a result, the executives are interested to hear about possible ways to improve this feature and bring more customers.

Mission

We develop a ranking model for recommending restaurants (other services as well) for groups with heterogeneous preferences. For customers, our model would return the restaurants that best match their preferences and maximize total utilities. For Yelp, our model is low-cost, flexible, and easy-to-implement.

Motivation

Let us imagine a group of friends who wanted to have an amazing restaurant dining experience in Las Vegas. Some of them were attracted by a French restaurant 30 min away. Some of them thought the nearby sushi place was good enough. Some of them were vegetarian. Some of them had kids with them……

They each opened Yelp and set the filters by own preferences. But it didn't help them decide because the app showed each of them very different results. As a result, our approach is to develop a data-driven recommendation system to aggregate the preferences and rank the matched restaurants.

Theory

Yelp helps customers maximize their utility by recommending restaurants to reduce information overload. However, when customers go out with others, it’s unlikely to find a restaurant that optimizes everyone’s utility. As a result, Yelp’s new objective is to help them find a restaurant that could potentially maximize their aggregated utility. Such utility should take both restaurant characteristics and group heterogeneity in preferences into account. The recommended restaurants will be those that best match with the group's expectations.

Model

Our model ranks the restaurants by the predicted utility score for the group at each possible restaurants. The utility scores are defined as $U_i = \frac{f(Ri)}{S(Ri,\ G)}$, where i indicates one restaurant, R_i is the preference attributes vector for each restaurant aggregated from previous reviewers, G is the preference attributes vector for each group aggregated from each member, $S(R_i, G)$ is the similarity score function that calculates the Euclidean distance of two vectors, and f(R) is the machine learning algorithm to predict baseline utility using the observable R.

In order to capture the features in R_i and G, we extract different information from the Yelp reviews. Also, we learned their predictive importance from Random Forest. Here are the 14 selected important features for the model:

- 'popular' (whether the restaurants have above-average ratings and review counts)
- 'Gender' (a proxy for female-friendly, defined as the percentage of female reviewers)
- 'Weekend' (a proxy for weekend-friendly, defined as the percentage of reviewers written at the weekends)
- 'Vegan' (a proxy for vegan-friendly, defined as the percentage of reviews that contain certain keywords)

- 'Kid' (a proxy for kids-friendly, defined as the percentage of reviews that contain certain keywords)
- 'Group' (a proxy for group-friendly, defined as the percentage of reviews that contain certain keywords)
- 'Ameri', 'Europ', 'SouAM', 'Asian', 'Indian', 'Chinese', 'MidEas', 'Drinks' (categories of food types)

Using R as predictors and rescaled customer ratings as the outcome, we trained 8 machine learning models to construct and find the optimal f(R). We decided to use Random Forest as it produces the lowest error calculated from cross-validation on the training set. Then we predicted the f(R) for almost 4k Las Vegas restaurants in the data.

When the group is making the decision together, G will be constructed automatically by linking each member's Yelp account, inferring individuals' preferences from previous activities, and aggregating the preferences. In addition, they can adjust their own preferences in the current mood. It will also enable each member to specify relative voting power, which would be helpful for parents to decide for the kids.

S(R, G) helps to reweight the baseline utility. It is a real number that measures the differences between the groups' ideal attributes with the restaurants' attributes. If the two vectors of attributes are close, S(R, G) will be small and scale up the baseline utility. If the two vectors of attributes are far away, S(R, G) will be large and scale down the baseline utility.

User Case

We go out to eat on the weekend, leaving from Flamingo Las Vegas (36.116111, -115.170556).

Each member's preferences are summarized as the following:

	popular	gender	weekend	vegan	kid	group	Ameri	Europ	SouAM	Asian	Indian	Chinese	MidEas	Drinks
Suresh (with kids)	1	0	1	1 (important)	1(important)	1	0	0	1	1	1	0	0	0
Li	1	0	1	0	0	1	1	1	0	1	1	1	0	0
Xi	1	1	1	0	0	1	0	1	0	0	0	1	1	1
Aggregated(G)	1	1/3	1	1	1	1	1/3	2/3	1/3	2/3	2/3	2/3	1/3	1/3

Figure 1: G (the preference attributes vector for each group aggregated from each member)

After calculating the similarity scores, we get U_i and returns the top 10 restaurants for this group:

	name	address	similarity	utility_reweighted	distance	walking_minutes
1	"Veggie House"	"5115 Spring Mountain Rd, Ste 203"	2.228397	24.772192	0.042045	0.502928
2	"Texas de Brazil"	"6533 Las Vegas Blvd S"	2.614818	20.173661	0.046224	0.507519
3	"Simply Pure by Chef Stacey Dougan"	"707 Fremont St, Ste 1310"	2.607002	20.137686	0.061099	0.842065
4	"Ronald's Donuts"	"4600 Spring Mountain Rd"	2.922497	18.943611	0.035252	0.441189
5	"Joyful House Chinese Cuisine"	"4601 Spring Mountain Rd"	2.669235	18.589845	0.034318	0.427196
6	"Del Frisco's Double Eagle Steak House"	"3925 Paradise Rd"	2.941861	18.444544	0.015233	0.169880
7	"Echo & Rig"	"440 S Rampart Blvd"	2.989858	18.382212	0.126205	1.661410
8	"Downtown Container Park"	"707 Fremont St"	2.532746	18.279102	0.061245	0.844358
9	"Fogo de Chão Brazilian Steakhouse"	"360 E Flamingo Rd"	2.874644	18.076145	0.014144	0.151941
10	"The Hush Puppy"	"7185 W Charleston Blvd"	2.769837	17.873966	0.090315	1.221522

Figure 2: Sample output of ranked restaurants