

Intelligent restaurant recommendation

Team member: Sarun Paisarnsrisomsuk, Tesia Shizume, Anthony Topper, Fangling Zhang

In the proposal, you should address the following issues:

What is exactly the function of your tool? That is, what will it do?

We use dataset from Yelp Dataset Challenge to build a recommendation system. It recommends restaurants to users based on their reviews of the restaurants that they have been to.

Why would we need such a tool and who would you expect to use it and benefit from it?

The system will provide a fast and accuracy recommendation. It will improve users' satisfaction and attract more users and restaurants to use its platform. Yelp, restaurants and users will all benefit from it.

Does this kind of tools already exist? If similar tools exist, how is your tool different from them?

Would people care about the difference? How hard is it to build such a tool? What is the challenge?

There are many recommendation systems, not only for Yelp, but also for other services, such as Amazon and Netflix. Most of the recommendation systems use collaborative filtering algorithms based on 'rating' variable. However, we believe that the reviews (long text) from users can provide more information than 'rating'. In our project, we will improve existing recommendation systems in three aspects:

1. We will use users' reviews to cluster users into different groups. We assume that users in the same group will have similar preference of restaurants. In our content-based collaborative filtering algorithm, we will recommend restaurants to a user **based on all users' rating in the same cluster**.
2. We will develop another model using Restricted Boltzmann Machine (RBM) to filter out a smaller number of restaurants. The model will be compared (and combined, if possible) with the model in aspect 1
3. We will use other attributes of restaurants like TakeOut, parking, categories to predict this users' rating, so that we can future adjust the restaurants' rank.

The challenge is the dataset is very large, it can take time and memory to complete this project. We may need to preprocess or filter out some users and restaurants.

How do you plan to build it? You should mention the data you will use and the core algorithm that you will implement.

1: How to cluster users: for each user, we combine his/her comments about all restaurants as one document, then use Vector Space Model to calculate the correlation between users. Next, we cluster the users.

2. How to transform and data: we set up a matrix with all restaurants as index, all users as columns, the rating score as value.

3. How to separate training data and evaluating data: For each user, we random sample, for example, 20% of the data as evaluating data, other as training data.

4: Besides RBM, we can also try auto-encoder. Ensemble all these model together.

5: How to use restaurant attributes to adjust rank: for each customer, use restaurant attributes value in his/her previous data to do regression, then put the target restaurant data in model to predict rating score, which can be used to adjust rank.

What existing resources can you use?

RBM pytorch package. NLTK package.

How will you demonstrate the usefulness of your tool?

F-measure and NDCG