**Intelligent restaurant recommendation**

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 Yelp Dataset to build an intelligent recommendation system. It recommends restaurants to customers more intelligently.

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

This system can develop the accuracy and speed of recommendation. It can develop customers satisfaction and attract more users and restaurants in its platform. Yelp, customers and restaurants 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 a bunch of people have set up Yelp recommendation systems. They use collaborative algorithms based on ‘rating’ variable. However, the reviews (long text) from customers provide more information than ‘rating’. In our project, we will try to improve the existed recommendation system in four aspects:

1. We mine user’s comments to cluster users. For each user, we combine his/her reviews about all restaurants as one document, then use tf-idf to figure out this user’s top 20 preference like price, drink, spicy, shrimp. Next, we use top words to set up some attributes like spicy-prefer, price-sensitive, drink-prefer. We cluster users with these attributes and other existed attributes in the dataset. In content-based collaborative algorithm, we recommend restaurants to a customer **based on people’s rating in the same cluster**.
2. We use deep learning algorithm restricted boltzmann machine to filter out 30 restaurants. It can be combined/compared with collaborative algorithm.
3. To adjust this 30 restaurants’ rank, we applied user’s comments (text) again. For each restaurant, we use NLTK and all its user’ reviews to sum the weighted **emotion score** of this user’s top 20 preference.
4. We can also 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 filter out some users and restaurants firstly.

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

In above answer**.**

**What existing resources can you use?**

RBM pytorch package. NLTK package.

**How will you demonstrate the usefulness of your tool?**

Cross validation.