Yelp Restaurant Recommender

Team 1: Bug Free
Team Member: Yujing Liu, Ying Xu, Yang Zheng

1 Introduction

Yelp, a platform that provides descriptions and reviews of various restaurants in the city, is an important tool in our daily lives. However, it does not to be able to recommend any restaurant for us automatically. Therefore, we study the characteristics of Yelp data, build a machine learning model (Collaborative Filtering) to perform rating prediction, and use the predicted ratings to recommend restaurants.

2 Methodology

2.1 Data gathering and processing

We use Yelp Open Dataset(https://www.yelp.com/dataset) to download the latest data, all data was in json format. Finally we use "Business", "Review", "User", "Tip" as our original data.

After unifying the case format of 'city' for business, we can see that there are 14,574 businesses in Philadelphia, which is the highest accounting for 9.7% of all data, so we choose Philadelphia as our research object. Then we went through their "categories" and identified those that contained "Restaurants" and were still open(3526). Beside reviews rating, tip is also an important reflect to show the user's feeling about the restaurant, so we used the "user_id" in the two datasets to filter out users who have commented or tipped on these restaurants. Finally, we save all these files into pickle format.

2.2 Model

Collaborative filtering is a way to predict the items that are most likely to be of current interest to users based on their past behavior on the platform (transactions, reviews, tips, etc.). It can be subdivided into user-based and item-based.

First, we create the binary diagram of user-item(ratings matrix) which reflects the rating given by users to restaurants. Then we use cosine similarity as a metric to measure the similarity of two restaurants or users, and make two predictions for item-based and user-based respectively.

At last, we assign weights to both by calculating the Root Mean Square Error (RMSE). After conducting several tests we find that 1:9 (item-based: user-based) yields the lowest RMSE in most cases, and gives our predicted star ratings in this ratio.

2.3 UI Design

Finally, we show users the Top 10 recommended restaurants. For visualization, we built a full-stack web application using Reactjs as the front-end framework and Flask as the back-end framework. In addition, we used Google Maps API to show the location of the final result on the map to make the interface more eye-catching.

3 Future Work

In the future, we can move further improving the running efficiency of our code and shortening the usage time. In addition, other machine learning methods, such as Content-based Filtering, can be used to perform more sophisticated processing of the data and get more reasonable recommendations.