Restaurant Recommendation

SI608 Project Proposal

Project members

Yanjun Chen, Jinyuan Chen, Yier Dong, Mingyu Li

Problem definition and motivation

Recommendation systems are being enforced to become more personalized nowadays. They need to be concerned about the users' social network and some other factors (Gomathi, et al., 2019). In this project, we are going to build a restaurant recommendation system based on users' social network as well as their ratings on restaurants. By doing so, restaurants will be recommended to the users whose close friends in their social network have highly rated the restaurants. Our main purpose is to allow close friends to have more common topics, namely their favorite restaurant. In addition, we believe that in most cases, people with close relationships are likely to have common tastes due to some objective conditions such as location, age and so on.

Basically, we will consider the following questions in the project to build the recommendation system:

1. Who is close to the user? How close are they?
2. Which restaurants are highly rated by the user's friends?
3. What network algorithm is going to be used? How to combine user social network and restaurant ratings?

By answering these questions, we will be able to build a restaurant recommendation system based on the users' social network.

Datasets that will be used and how these data will be collected

Data Source:

The primary dataset that we will be using is from Yelp dataset, which can be downloaded from https://www.yelp.com/dataset. The dataset consists of 5 JSON files: business.json, review.json, user.json, checkin.json and tip.json. Examples for the data fields of the 5 JSON files can be previewed on https://www.yelp.com/dataset/documentation/main.

Data Preprocessing:

As the size of the original dataset is relatively large (8.65GB), we plan to use only a subset of them by dropping some irrelevant features to our research interest for this project. For example, the review.json file (4.97 GB) contains text reviews which take a lot storage

space but might not be helpful if we are not going to conduct NLP analysis on these texts, so we will drop this feature and keep other features like stars, user_id, business_id etc.

Data Sampling:

Getting a subset of the dataset by sampling might also be helpful in reducing the data size, but as user.json contains a “friends” field which is an array of the user's friend as user_ids, we cannot just simply sample the data entries randomly, which will very likely to make some friends of a user missing in the subset data. We will try data sampling if our team finds a better way to do this during the project.

Analysis/methodology/system/experiment plan

To find out those having close relationships with a specific user, we will build a user network. Nodes are the users and edges are the relationships. Hopefully, the edge should be weighted based on the user information. Currently, we do not expect that the relationship is directed. In addition, we will have a network showing how users rate the restaurants. Nodes are users and restaurants, while edges are ratings. The networks will definitely help us find out whose highly-rated restaurants should be recommended to the user in priority.

Brief timeline

By the end of September: Clean and pre-process the data, explore different methods and models

By the end of October: Build the network, implement the models, run experiments with the models

By the end of November: Collect the results and evaluate the performance, do network analysis based on the results

By the end of the Semester: Finish all the experiments, write the final report

Backup plan

We will switch to Anime Recommendations Database (https://www.kaggle.com/datasets/CooperUnion/anime-recommendations-database) and build an anime recommendation system based on this dataset. Because this dataset is really small (26MB) and really simple (7 columns), we use this as our backup plan if the yelp data is too huge for our computers.

Reference

R. M. Gomathi, P. Ajitha, G. H. S. Krishna and I. H. Pranay, "Restaurant Recommendation System for User Preference and Services Based on Rating and Amenities," 2019

International Conference on Computational Intelligence in Data Science (ICCIDS), 2019, pp. 1-6, doi: 10.1109/ICCIDS.2019.8862048.