

Dish recommend

based on yelp

Problem

- What topics customer have in reviews?
- How can we make a better recommendation for users?
- What dishes the users like in different cuisine?
- Can we predict hygiene situation for restaurants through reviews?

Approach

- Using python json package to collect the data.
- Using tf-idf, document term frequency and other method to preprocess the data.
- Using the LDA/PLSA algorithms to get the topics from reviews.
- Using K-means and other algorithms to cluster the data.
- Using similarity matrix to see the different kind of cuisine.
- Using sentiment analysis to mine the favorite dishes.
- Using the naïve Bayes, random forest and other algorithms to make a prediction.
- Using off-shell analysis tools to process data.

Data

- The rate of restaurant.
- The review of restaurant.
- The business information data.
- We plan to get up to 1G data for this project.
- And we will get the data from Yelp API.
- We may manually label some data when we want to find more dishes from reviews.

Timeline

- To collect review from yelp using yelp API before 10/25.
- Data preprocessing for retrieving the restaurant reviews before 11/01.
- Clustering the subsets and analyze the topics before 11/08.
- To recommend users for their favor dishes restaurant before 11/15.
- Predicting hygiene situation for restaurants through reviews before 11/22.