Recommendation System Based on Yelp Dataset

Group Member: Juan Lin, Yan Li, Yichen Wang, Xu Han

Motivation

1. Users are being presented with tremendous ranges of choices.
2. Sellers are being faced with the challenge of personalizing their advertising efforts.
3. Recommend items or products to a collection of current users or potential users

Data Description

1. 5.2 million reviews, 174 thousand businesses, and over 1.2 million business attributes.
2. uncompressed SQL file containing 7.55 Gbs.
3. store data in an Amazon RDS in AWS

Proposed Work

- **Proposed work:** We will build utility matrices with users' ratings data and use several techniques to predict recommendation results. The techniques we use include machine learning techniques(feature engineering and model training), content-based algorithm and collaborative filtering algorithm.
- **Evaluation:** We plan to divide the dataset as training part and validation part. The ratio is 4:1. It will be used for evaluating machine learning (precision, recall, etc) and recommendation system.
- **Timeline:** **Mar.16** - Data preprocessing and initial feature engineering; **Apr.6** - Feature engineering and model training; **Apr.20** - Evaluation and website building; **May.4** - Other RS algorithms implementation and report writing