Team information: Ravneet Singh Sidhu & Harsha Vardhan Gorla

Objective and Overview of project:

Develop a web application that will find a friend (Yelp user) nearby to the user location based on the common interests. We compare the reviews and the most frequent type of restaurants or shopping malls visited by the users to get common interests of both the users. If both the users want to meet up and have food together the website will suggest a meeting place (based on their reviews and most visited restaurants) for them to hangout.

Data Mining tasks:

1. *Sentiment analysis* refers to various methods of examining and processing data in order to identify a subjective response, usually a general mood or a group's opinions about a specific topic. Sentiment data is often derived from social media services and similar user-generated content, such as reviews, comments, and discussion groups. The data sets thus tend to grow large enough to be considered "big data."[1]
2. Map-Reduce to analyze the user and business data.
3. Predictive algorithms like Naïve Bayes classifiers and Clustering techniques.

Plan:

A web application that allows Yelp user to login and find out friends who have similar interests and live nearby. On the consent of both users the application will suggest a meeting place based on their previous reviews and locations.

Challenges:

We have many challenges which arise from the dataset. First we need to clean the data as different items across different restaurants are written differently. Secondly application of sentimental analysis on user reviews and then comparing the common interests among list of users based on likes, common friends etc. Thirdly the application performance due to large dataset and need of an efficient algorithm will be of great importance.

Address Challenges:

We will try to model data based on cuisine, type of restaurant and location of restaurant to look for similar patterns. This would require intelligent use of data and making a decision based model to give concrete results. In order to scale the application at increase the application performance we will try to implement the logic using Amazon Web Service. The use of hadoop with map-reduce may also help us in achieving faster results.

Evaluate Efficacy:

The efficacy would be determined by predictive matches the application delivers using a single user data based on factors like location, reviews etc. However the effectiveness of using food and reviews as a primary factor in predicting human behaviors such as long term compatibility would not be covered in our application. It may be considered as a future area of study.

Task Assigments:

Data cleaning, predictive algorithms, Website: Ravneet

Sentimental analysis, Map-Reduce tasks, Cloud: Harsha

References:

1. http://docs.aws.amazon.com/gettingstarted/latest/emr/getting-started-emr-sentiment-tutorial.html - AWS