Project Progress

Recommender System for Yelp Users

Team Members: Hardik Manoj Katariya 1001030751

Current Status

Project Objectives and Tasks:

The objective of the project is to build a recommender system which would help predict the rating a user might give to a given item (business in our case) and also recommend a particular item (business) to the user.

This would require the following tasks:

- Find out all the similar users.
- Find out all the similar businesses.
- Perform sentiment analysis for businesses based on the user reviews for the business. Also understand the type of businesses the user likes based on the reviews the user has given.
- Build a user-user as well as item-item collaborative recommender system.

Project Progress:

Currently I am not finished with the things such that I could present them. I would like to describe the work I have started for the tasks.

- Sentiment Analysis: I have started work on finding various algorithms and API's that would help me classify the reviews as positive negative or neutral. I found an API in python called 'TextBlob' which helps in classification. Initially I found the accuracy to be around 50% but later realized that it could not handle neutral data efficiently. So now I am trying to build a Bayesian classifier. I plan to use the data given at http://www.sananalytics.com/lab/twitter-sentiment/ in order to train the classifier as well as I have dataset used by 'sentiment140', a sentiment analysis application, to train its classifier. If the accuracy of the classifier is good I would use it or else I plan to proceed with 'TextBlob'.
- Collaborative Filtering: I was new to this concept so I started research on the topic. Currently I am doing a course called 'Introduction to Recommender System' on courser.com. I would be finishing the course in a week. Once I am done with it I would start building the recommender system.
- Website: I plan to build a website which would have this recommender system. Since I am finding the similar users and similar businesses, I plan to show some clustering as well using D3.js (If time permits). Highly unlikely due to lack of time and resource, but I could also present which businesses are more successful in a particular location (I would definitely try).

Also, one of the major issues I face is that my laptop is not performing well. So I plan to do the project on cloud using AWS EC2 instances. I plan to do the computation and analysis using Hadoop Map Reduce method. I also plan to host the website on AWS.

Tasks to be accomplished:

As mentioned above, none of the tasks has been fully completed. So all the tasks will be completed before the final deliverable.

References:

- http://www.sananalytics.com/lab/twitter-sentiment/
- http://www.sentiment140.com/
- http://help.sentiment140.com/for-students/
- https://www.coursera.org/learn/recommender-systems
- http://textblob.readthedocs.org/en/dev/