

AC209a Milestone 5: Data Science with User Ratings and Reviews

Andrew Ross, Sophie Hilgard, Reiko Nishihara, Nick Hoernle

November 27, 2016

Proposal of Future Work

For Baseline Models in Milestone 4, we have implemented simple averaging as well as basic content-based recommendations and collaborative filtering models. Going forward, we hope to build more sophisticated models within these classes. In particular, we will explore creating a hybrid collaborative filtering/matrix factorization model or a model which more heavily weights recent ratings.

We also built a network-based recommendation system based roughly on the PageRank algorithm. This model finds the most ‘connected’ business, where path lengths between businesses are weighted inversely to how much a user liked both of the restaurants. To find the next recommendation for any given user, we search the graph built off of similar users and return the ‘closest’ business the user has not yet been to.

We have also explored a number of other models for predicting business success rather than focusing only on recommendation systems. In particular, we have spent time focusing on other aspects within the yelp challenge that ask you to predict business success based on location or investigate seasonal and otherwise temporal trends in business popularity. Location based models have focused largely on clustering methods, as neighborhood is a better predictor than absolute latitude and longitude values. Trend analysis will use regression models over time to analyze if businesses are cyclically popular at any given lag or to find points in time when a given business was comparatively ‘hot’ within its lifetime.