

AC209a Milestone 5: Data Science with User Ratings and Reviews

Andrew Ross, Sophie Hilgard, Reiko Nishihara, Nick Hoernle

November 28, 2016

Proposal of Future Work

In Milestone 4, we evaluated the performance of a number of baseline recommendation models (simple averaging, matrix factorization, popularity ranking, and PageRank) on restaurants and bars in the Yelp dataset. Going forward, and following some of the techniques used in [1], we want to see if we can surpass those benchmarks by blending models, computing time-sensitive user and business baseline ratings, and by exploring other models like collaborative filtering. We're also interested in exploring models that exploit more subtle time and location-based effects, as well as continuing to explore the data itself.

Blending

One important feature of the Netflix Grand Prize solution, and an increasingly talked about technique in general, is blending predictions from different models to achieve a lower overall error. Concretely, this usually means taking a linear combination of the individual model predictions, where the weights are computed per-example using another model (gradient-boosted decision trees are a popular option [1, 2]). It is also common to train separate blend weight models for different subsets of users [2], and we've also seen in homeworks that there are also majority vote or selection-based methods for combining predictions, which may be more appropriate for classification tasks (depending on whether you treat predicting ratings as classification or regression – both perspectives seem to have merit). We plan to explore at least a few of these methods.

Time-sensitive baselines

Another important feature of [1] is the use of time-changing baseline predictors, as well as a frequency baseline that accounts for the number of ratings a user makes in a day, which was important in the Netflix dataset. As we saw in our data exploration, it is also important in the Yelp dataset; 1- and 5-star reviews mean something different coming from a user rating many businesses per day than they do from a user rating just one. We will attempt to evaluate how including each particular term in our the baseline predictors affects our RMSE.

Even more time-sensitive baselines

In many ways, time is much more of a factor for Yelpers seeking food or drinks than it is for Netflix users picking their next show. Although public opinion about a show or movie sometimes changes over time, the actual quality of a restaurant's food can change much more rapidly. This can make past restaurant rating data less reliable for predicting future ratings. We're interested in seeing if we can reduce RMSE by weighting recent ratings more heavily when computing baselines, or perhaps trying to detect anomalies. To evaluate performance, however, we may have to test our performance only against recent ratings in our test set.

Another factor we noticed was that time of day and day of week had a strong correlation with ratings for certain types of businesses. This was especially true for bars and "Nightlife" businesses (shown below), and location also came into play. The plot below shows the ratings that are above the mean in red and the ratings that are below the mean in blue. The average number of ratings per day of the week is shown in the right side vertical axis and the number of ratings on any one day of the year is shown in the axis above the plot. We clearly see that the reviews that are given on Sundays and Mondays tends to be more positive than the reviews given on the other days of the week.

Weekly or seasonal effects were not something discussed at all in [1], perhaps because they aren't as meaningful for movies (It's a Wonderful Life and Bad Santa aside), but we're interested in seeing if we can use them to improve our predictions for restaurants and bars a bit further.

Location-based analysis and going beyond rating RMSE

We have also explored a number of other models for predicting business success rather than focusing only on recommendation systems. For example, we can use the Yelp dataset to measure business popularity (using review counts or checkin volume as well as average rating), or even whether and when a business eventually closes. We've spent time investigating location-based, seasonal, and other temporal trends in these business metrics. 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.

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.

References

- [1] Yehuda Koren, [The BellKor Solution to the Netflix Grand Prize](#)

- [2] Michael Jahrer, Andreas Tscher, and Robert Legenstein, [Combining Predictions for an accurate Recommender System](#)