Improving Yelp Restaurant Recommendations

Gil Rosen, Tuan Nguyen

Background

Abundant amounts of textual data available from reviews on Yelp have been used extensively for recommendations. However, major challenges remain in generating personalized recommendation. One of these challenges involves taking into account business' ratings and customers' preferences when matching customers to products that they would like.

Previous work by Jack Linshi demonstrated effectiveness of the common hypothesis that star rating is justified by corresponding review. In particular, running LDA model on corpuses conditioned on the star rating levels yields much more informative semantic and sentiment topical aspects.

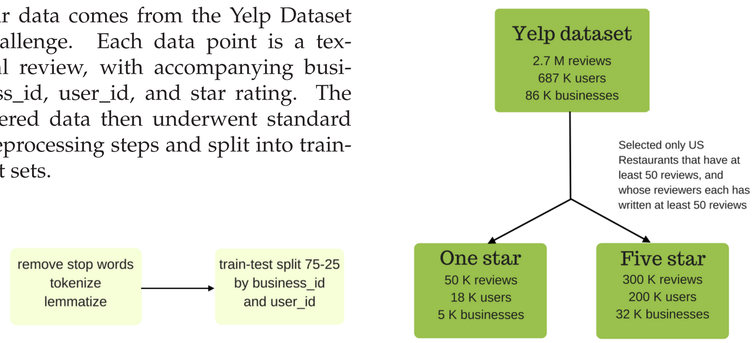
This project investigates how to identify users preferences based on their existing reviews conditional on star ratings and take into account these preferences when making recommendations of restaurants to be relevant for this user.

Problem Statement

Current reviews on Yelp are predicated on an overall score for a given establishment with a very rough "star" system and a textual review. Star ratings do not incorporate the abundance of valuable information in the textual review itself. Our objectives are to evaluate whether topics generated by LDA conditioned on star ratings can be used to automatically generate user preference profiles and restaurant characteristic profiles that could be useful in matching customers to restaurants, as well as to build a K-means model which may accurately predict a given users rating for establishments.

Data & Preprocessing

Our data comes from the Yelp Dataset Challenge. Each data point is a textual review, with accompanying business_id, user_id, and star rating. The filtered data then underwent standard preprocessing steps and split into train-test sets.



Methods & Models

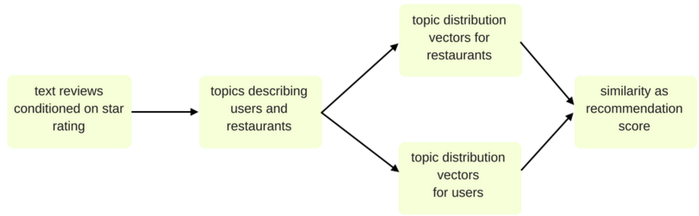
K-Means

- Samples are defined as the set of reviews across all restaurants per user; each user represents a new sampling.
- We created clusters to denote users with similar tastes - the assumption being that those exhibiting vector review profiles similar to one another would rate similarly, both positively and negatively, establishments that they had not visited
- For the test set we also removed 1/4 of the business reviews to test if they would be accurately predicted by the model.

Methods & Models (Continued)

LDA

- LDA is used to discover topics hidden in the text reviews corpus, conditioned by their star ratings.
- A restaurant is described by a vector, average of the topic distribution vectors produced by the LDA model for the restaurant's reviews.
- Similarly, a user is described by a vector, average of the topic distribution vectors produced by the LDA model for his/her reviews.
- Similarity score between a user and a restaurant can determine whether a user would like the restaurant.



Experimental Results

Models training

The number of topics k were determined experimentally. We trained LDA models using $k = 20, 30$, and 40, each with 30 passes through the corpus.

To the right is a perplexity plot for training with $k = 30$ model. Perplexity decreases and stabilizes around 20 passes. Performance for LDA is evaluated using Mean Square Error between a user's rating from the test set and the predicted rating. $k = 30$ performed well, while $k = 40$ which took quite longer to train, only improved marginally.

The number of clusters C for our K Means model was also determined experimentally with 50 clusters providing the most value for processing time. Performance was measured by the proportion of withheld restaurants from a user in the test group which their cluster rated high/low, that the user indeed rated similarly.

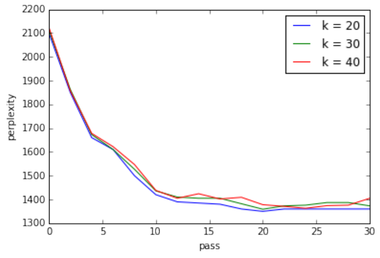


Figure 1: Perplexity plot

Results summary

Number of topics	Optimal leading α	Optimal leading η	MSE
20	0.029	0.032	1.9365
30	0.031	0.029	1.4473
40	0.027	0.033	1.3208

The topics generated by the LDA models indicate quite well features that can differentiate restaurants and users. Sample topics from our results describe Italian, Japanese, French, South American cuisines, among others.

Results (continued)

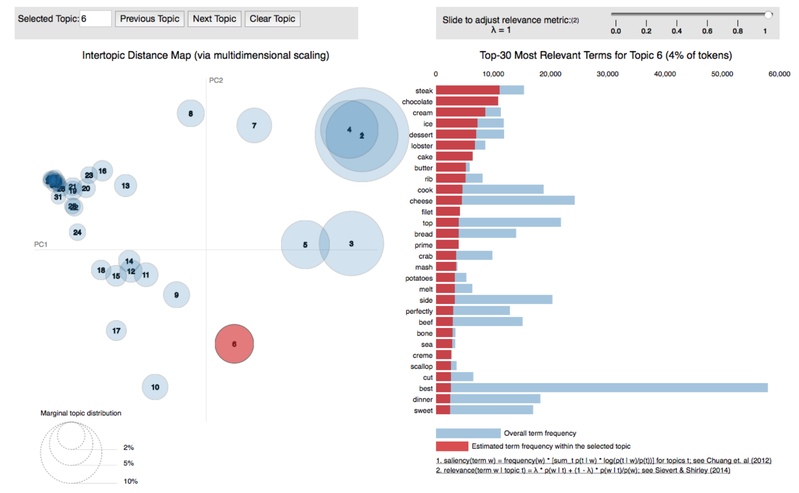


Figure 2: Sample topic from LDA results

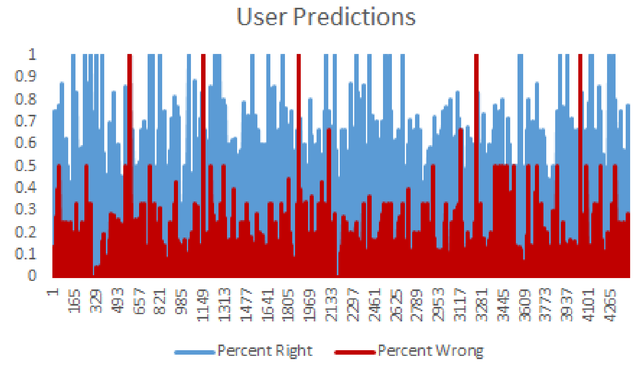


Figure 3: K-means results

Likewise, our K means prediction succeeded in predicting 24 percent of those establishments users later reviewed similarly. More critically, it only exhibited 3.4 percent incorrect predictions - establishments that users rated differently from their group.

Conclusions

- LDA models trained on reviews conditioned on star ratings generate topics that can accurately describe restaurants and users. $k = 40$ topics gives reasonably good performance, indicated by MSE of 1.3208.
- Future investigation should explore connection between LDA and word2vec.
- K means provides a valuable measure of predicting user preferences for establishments with little error.
- Given the much larger set of available establishments (dimensions) to those reviewed by any given user, this provides ample opportunity for accurate recommendations.