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.

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.

RESULTS (CONTINUED)

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 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.

While 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.

Figure 1: Perplexity plot
Figure 2: Sample topic from LDA results
Figure 3: K-means results

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 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.

While 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.

Figure 1: Perplexity plot
Figure 2: Sample topic from LDA results
Figure 3: K-means results

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.