RESOURCES

yelp DATASET CHALLENGE

vowpal wabbit

FOOD WITH FRIENDS

YELP GROUP RECOMMENDATION SYSTEM

STEPHEN COGNETTA
PRANAV GOKHALE
EUGENE LEE
MATHILDA LLOYD

ABSTRACT

When going out with friends or perhaps a group you've never met, how do you select a restaurant that everyone will like?

We analyze data from Yelp Dataset Challenge 2014 which contains comprehensive business rating data for Phoenix, Arizona.

We address two questions:

- Generating a list of restaurant recommendations for a given group of users, using collaborative filtering and introducing a social value function, which combines users' interests in a way to satisfy the whole group
- Finding a group of compatible users given their previous rating history, using similarity measures

We successfully predict ratings with an RMSE better than a previous Yelp RecSys Challenge's best RMSE, and we have promising validation results.

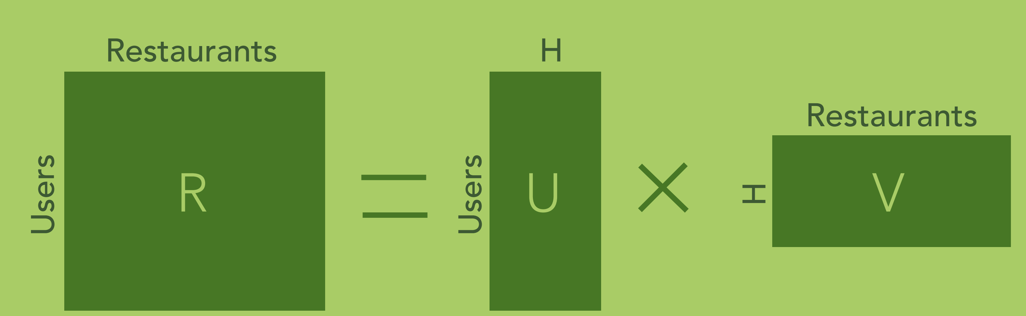
DATA

Notation

- U and B are the total number of users and businesses, respectively.
- Ratings are contained in matrix R, with $R(u,b)$ representing user u's rating of business b.
- $S \in U$ denotes a specific set of users, and $S_u \in B$ denotes the set of businesses user u has rated.

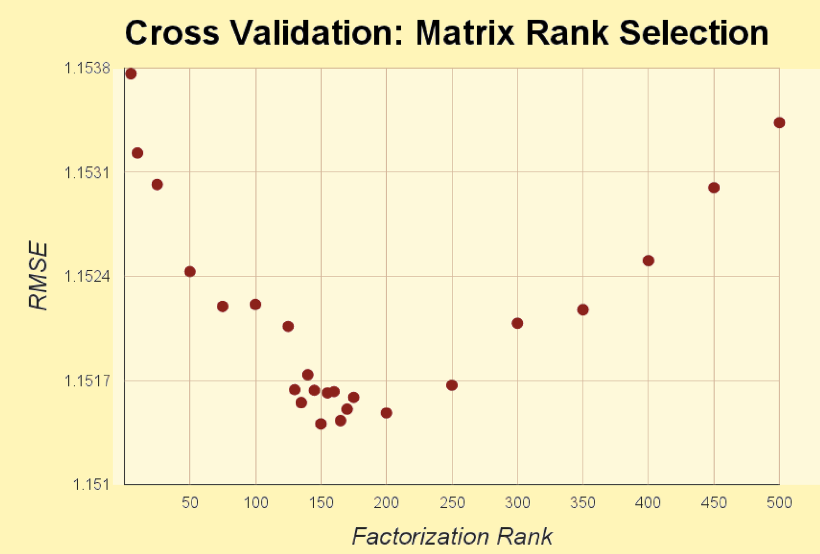
Postprocessing

- We filter business to only include businesses with the tags "Food" and "Restaurant" since this is our project focus.
- We optionally normalize users' reviews to ensure they have the same underlying meaning. By normalizing each member's ratings to the average group rating, we scale the ratings to accurately reflect members' utilities.



Introduction

- We apply collaborative filtering using industry standard tools.
- To use matrix decomposition, we cross-validate to achieve a final RMSE of 1.15, better than previous Yelp Challenges.



Social Value Functions

To combine users into groups, we apply these social value functions:

MOST HAPPINESS

The person who likes the restaurant the most will determine group sentiment.

$$g_{mh,S}(b) = \max_{u \in S}(R(u,b))$$

EXPERT

Users who have rated more restaurants will have more influence over group sentiment.

$$g_{exp,S}(b) = \frac{\sum_{u \in S} R(u,b)|S_u|}{\sum_{u \in S} |S_u|}$$

LEAST MISERY

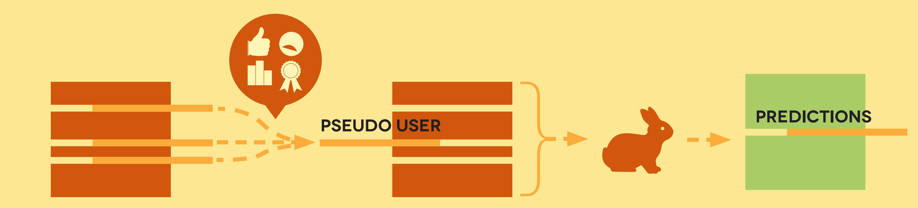
The person who likes the restaurant the least will determine group sentiment.

$$g_{lm,S}(b) = \min_{u \in S}(R(u,b))$$

AVERAGE

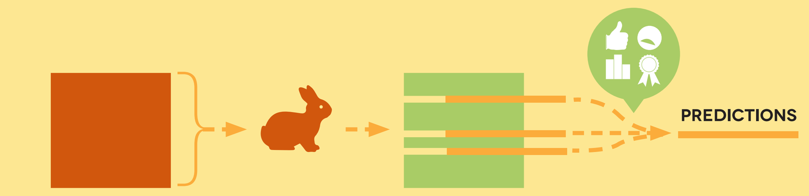
The group rating for each business is the average of each member's rating for the business.

$$g_{ave,S}(b) = \frac{\sum_{u \in S} R(u,b)}{|S|}$$



PSEUDO-USER APPROACH

We create a "pseudo-user" to represent the tastes of the individual group members and run the individual recommendation system on the pseudo-user.



POST-MERGE APPROACH

We run our individual recommendation system on the unmodified R, generating predicted ratings for all group members, then apply the social value function on these predicted member rating vectors.

Evaluation functions

We remove a set of ratings for the same restaurant b from R, then generate our group recommendation vector for the set of users S whose reviews we removed.

ACCURACY

This statistic measures the difference between the average true ratings of the group members and the predicted group rating.

$$\text{Accuracy} = |g_S(b) - \frac{\sum_{u \in S} R(u,b)}{|S|}|$$

FAIRNESS

We measure the increase in variance as the difference between the actual reviews' deviation from the true group average rating, and the actual reviews' deviation from the predicted group rating.

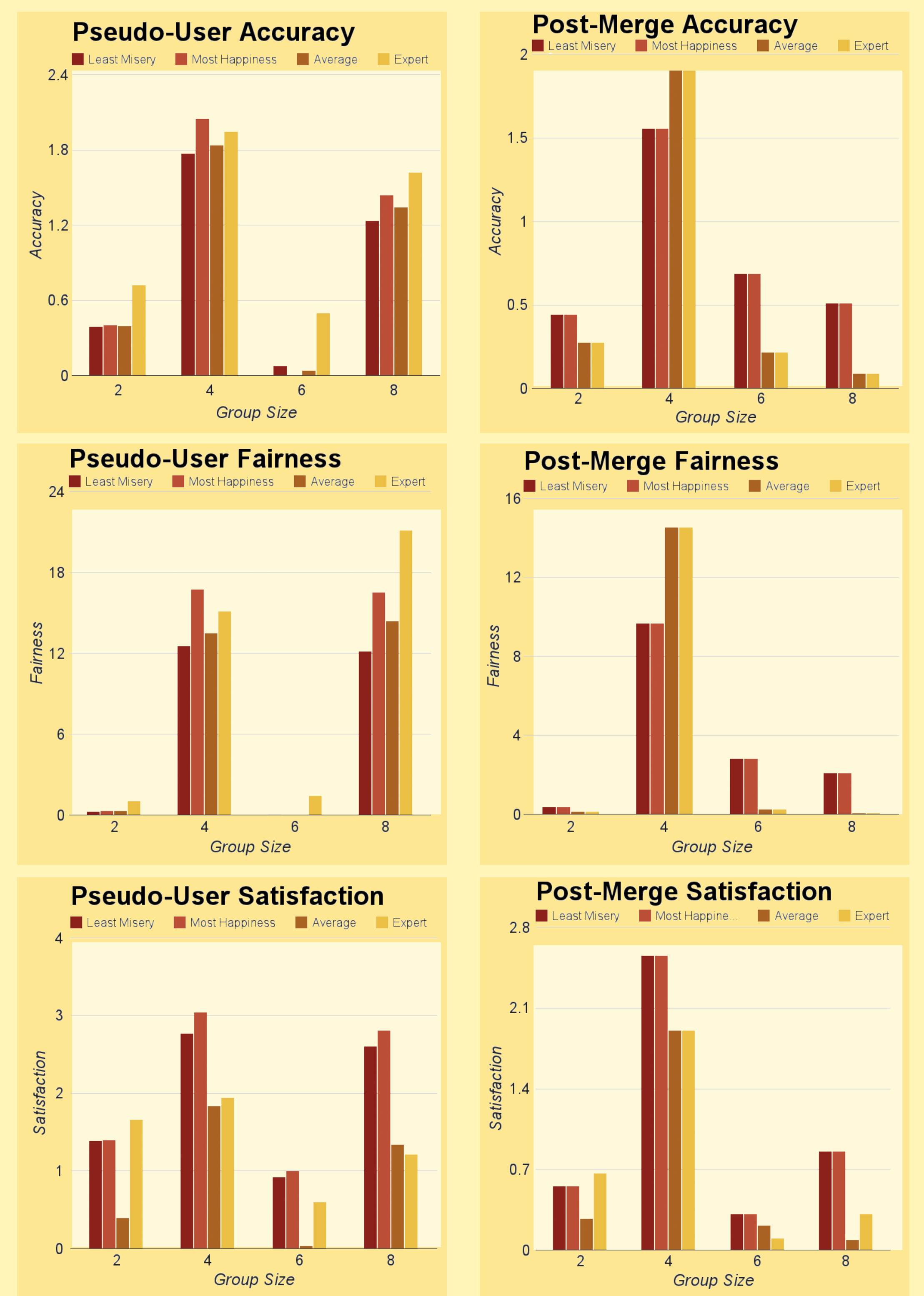
$$\text{Fairness} = |\sum_{u \in S}(R(u,b) - g_S(b))^2 - \sum_{u \in S}(R(u,b) - \frac{\sum_{u \in S} R(u,b)}{|S|})^2|$$

SATISFACTION

Assuming our social value function is valid, this metric calculates the difference between the actual group utility and the predicted group rating.

$$\text{Satisfaction} = |f_{sv}(R(S,b)) - g_S(b)|$$

RESTAURANT PREDICTION



Analysis

- Pseudo-user approach is more computationally intensive than the post-merge approach.
- The expert social value function generally performs poorly.
- Least misery and most happiness have similar evaluation statistics due to their similar social value function schemes, but have poor satisfaction metrics because they trend towards extreme values.
- Overall, the data suggest that our approaches are relatively consistent and offer nearly identical recommendations.

GROUP RECOMMENDATION

Introduction

Our goal is to identify users with similar preferences for restaurants, and recommend groups based on those preferences.

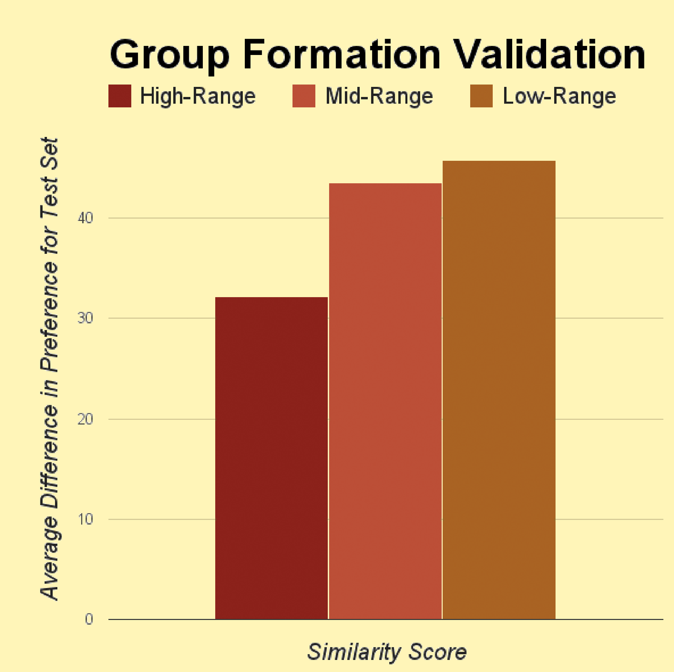
Method

- We calculate their similarity using a cosine similarity function. For users u and u' that have overlapping ratings vectors $R'(u)$ and $R'(u')$ respectively, we calculate their similarity $\text{sim}(u,u')$.
- We then sort the set of neighbors over their calculated similarity score, to generate a list of users with preferences similar to those of the given user.

$$\text{sim}(u,u) = \frac{R(u) \cdot R(u)}{\|R(u)\| * \|R(u)\|}$$

Validation

- For a random given user's rating vector, we remove the rating for randomly-chosen restaurant b. After running the group formation system, we calculate the difference between the neighbors' actual rating of b and the given user's rating of b
- We expect this difference to increase as our similarity score for the user decreases.



Analysis

- Users with high similarity to the selected user have similar preferences to the removed test restaurant.
- Therefore our similarity measure can accurately judge similarity for new restaurant selections.

CONCLUSION

We have successfully developed solutions for both of our original problems. We have created a group recommendation system using collaborative filtering techniques. We have also generated a scheme to find users with similar preferences to a given individual.

Our group sincerely appreciates the support of Dr. Laurent Charlin and Prof David Blei.

References

Yelp Dataset Challenge citation: http://www.yelp.com/dataset_challenge/
Vowpal Wabbit citation to documentation: http://hunch.net/~vw/
Similarity calculations reference: http://www.cs.carleton.edu/cs_comps/0607/recommend/recommender/itembased.html
http://link.springer.com/chapter/10.1007/978-3-540-72079-9_9#page-2
http://dl.acm.org/citation.cfm?id=963772
http://files.grouplens.org/papers/poly-camera-final.pdf