Effects of Social Influences on Culinary Preferences

Dhawal Joharapurkar
Department of Computer Science
University of California, Santa Cruz
Santa Cruz, CA 95064
dhawal@ucsc.edu

Varsha Joshi
Department of Computer Science
University of California, Santa Cruz
Santa Cruz, CA 95064
vajoshi@ucsc.edu

Abstract

Higgs et al., [HT16] hypothesize that an individuals' eating behaviour is strongly influenced by social context. In this work, we seek to model the influence of a friend network on the culinary choices of individuals. We predict a person's favorite cuisine from social influences from their network and evaluate our model by comparing our predictions to the gold labels that are extracted from the text of the reviews. We use the data from the "Yelp Dataset Challenge" dataset [yel].

1 Project Objective

Higgs et al., [HT16] talk about the effects of social influences on eating are powerful and pervasive and may play a role in the development and maintenance of obesity. We also feel that social influences affect culinary preferences and is an interesting topic to study.

While collaborative filtering for restaurant recommendation based on a person's social network is a well studied topic [LKJJ06] [PPC08], we look to generalize the task to be able to identify a person's favorite cuisine which therefore could be a front to restaurant recommendations. We believe that in doing so, we are allowing for recommendations to be more flexible.

We believe that there is a mix of personal preference and network influence in determining a person's cuisine predilection. Using the Yelp dataset allows us to mine both aspects of the problem. For personal preference - we can extract a person's check ins into different restaurants in the hope that a person is more likely to check into restaurants that serve a certain cuisine more than others if he/she tends to enjoy that kind of food. For network influence, we look to use the Yelp friendship network, which is essentially a Facebook friendship network, since when people sign onto Yelp with their Facebook credentials, they are automatically added as friends on Yelp. Culinary influences tend to flow in a friendship network, since eating is not a solitary event, and people generally tend to go out to eat and drink in groups, thereby influencing each other's tastes and preferences.

In this project, we look to label users on Yelp with a "cuisine" - which can be thought of a cross between their personal culinary preference and the influence of the food choices of their friend network. This is a multi-class classification problem, since a person can have more than one favorite cuisines. To this effect, we followed these steps:

- Extract users with a predilection for cuisines by mining their reviews for keywords such as "favorite cuisine", "love X cuisine", etc.
- For this set of users, extract their friendship graph
- Using PSL, model the cuisine dispersion through the network via collective classification

30th Conference on Neural Information Processing Systems (NIPS 2016), Barcelona, Spain.

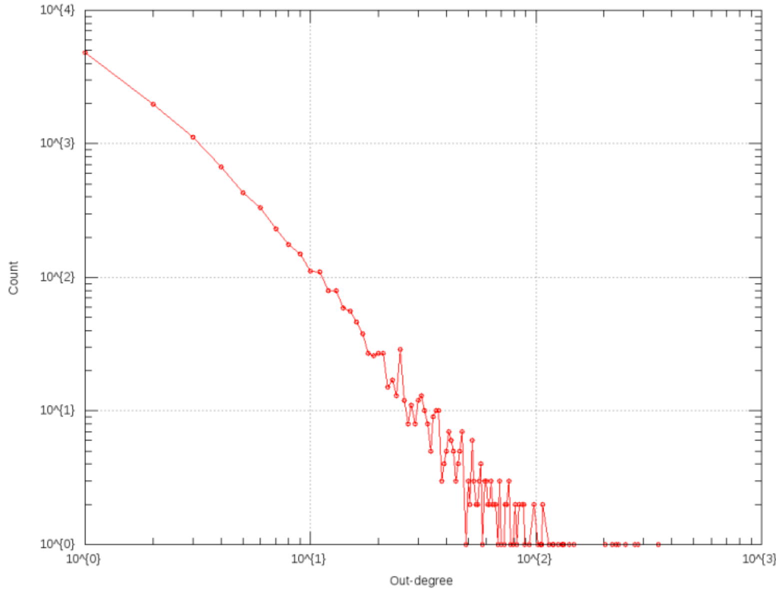

Figure 1: Degree Distribution of Social Subgraph

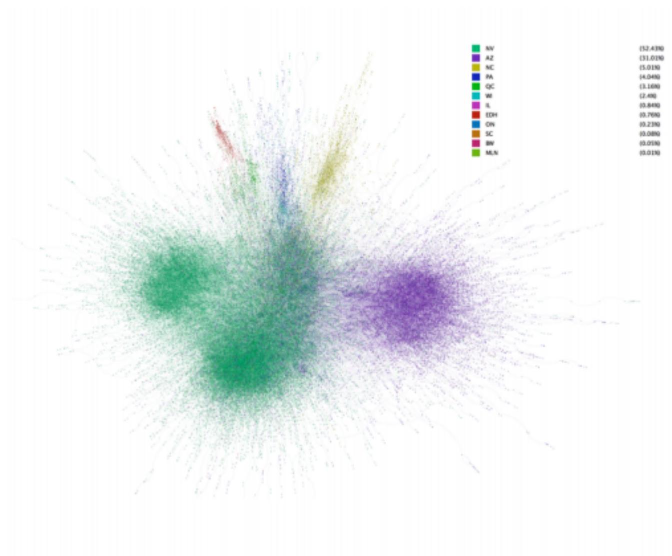

Figure 2: Yelp Friendship Network Colored by Location

2 Dataset

The Yelp dataset provides business information across 10 different cities and the data is divided into five different JSON files. The structured dataset provides information about businesses such as location, neighbourhood and open times. It also lists various attributes for users such as user friends, fans, useful votes earned, etc. The dataset contains a review file which is a collection of reviews written by users for all the businesses. The reviews along with user's opinion also has attributes such as star rating given to the business.

Reviews in the dataset were used to extract favourite cuisine information for users. There are over 4 million reviews in the file and users follow a near power law distribution with reviews [CW12]. It was observed that users mention phrases such as 'I love X cuisine', 'X is my favourite', etc. i n their reviews. We parsed these phrases to map favourite cuisine with an individual user. We found 31 distinct cuisines in the reviews file. We used users' friends list to extract the friends network on Yelp.

The established network has 14106 unique users with favourite cuisine information. The friend list ranges from 0-715 average number of friends being 15. We used this data from these users to build Probabilistic Soft Logic (PSL) [KBB+12] observation files.

3 Baseline

To set a baseline for the performance we implemented SVM classifier to predict cuisine labels for individual users. We used SVM (C=1.0, kernel='rbf', gamma='auto') classifier from scikit learn library.

The baseline model uses the assumption that a users visiting similar restaurants and giving similar ratings to these restaurants would have similar cuisine preferences. The feature vector for the baseline was created by combining all the restaurants visited by users in the friends network. For individual users, feature vector was created by calculating average rating given to each restaurant.

To enhance the model further we added key cuisine words mentioned by the user in his/her reviews. The feature vector was appended with cuisine occurrence count for each cuisine. This improved the performance of the model significantly.

Features	Accuracy
Average rating for restaurants	12.08 %
Average rating for restaurants + cuisine occurrence in reviews	31.21 %

Table 1: Baseline Evaluation

The iid (independent and identically distributed) assumption makes the model weak which is evident by the performance. As explained above culinary choices of users can be better predicted using inter-relations between friends.

4 PSL Model

4.1 Predicates

We list the predicates with their arities and variables:

- favoriteCuisine(P,C): person P's favorite cuisine C
- Friend(P1,P2): person P1 is friends with person P2
- socialInfluenceOnCuisine(P,C): person P is socially influenced towards cuisine C
- usefulUser(P): the person P is a useful user on Yelp. We filtered the users to keep the people who have more useful votes than the average number of useful votes.
- funnyUser(P): the person P is a funny user on Yelp. We filtered the users to keep the people who have more funny votes than the average number of funny votes.
- coolUser(P): the person P is a cool user on Yelp. We filtered the users to keep the people who have more cool votes than the average number of cool votes.

4.2 Model

The current model runs following rules to compute a user's friend networks influence on various cuisines. By running the rules below we get truth values for both Social Influence on Cuisine and favourite cuisine for a user.

- **5**: favoriteCuisine(P,C)
- **10**: (Friend(P1,P2) & favoriteCuisine(P1,C)) » socialInfluenceOnCuisine(P2,C)

Popularity Rules	Funct. Constraints	Accuracy	Precision	Recall	F1 Score
UsefulUser & coolUser & funnyUser	NA	36.70 %	Thai: 0.2000 Chinese: 0.3000 Mexican: 0.5172	Thai: 0.9000 Chinese: 0.6923 Mexican: 0.5000	Thai: 1.8461 Chinese: 2.7000 Mexican: 1.5000
UsefulUser & coolUser & funnyUser	Yes	49.78 %	Thai: 1.0000 Chinese: 0.9230 Mexican: 0.8064	Thai: 0.6153 Chinese: 0.4675 Mexican: 0.4716	Thai: 1.8461 Chinese: 1.4025 Mexican: 1.4150
UsefulUser & coolUser	Yes	50.63 %	Thai: 1.0000 Chinese: 0.9250 Mexican: 0.8064	Thai: 0.6153 Chinese: 0.4805 Mexican: 0.4629	Thai: 1.8461 Chinese: 1.4415 Mexican: 1.3888
NA	Yes	51.47 %	Thai: 1.0000 Chinese: 0.9230 Mexican: 0.8787	Thai: 0.6153 Chinese: 0.4800 Mexican: 0.4328	Thai: 1.8461 Chinese: 1.4400 Mexican: 0.2985

Table 2: Performance of PSL

- **10**: (socialInfluenceOnCuisine(P1,C) & Friend(P1, P2)) » socialInfluenceOnCuisine(P2,C)
- **10**: (socialInfluenceOnCuisine(P,C)) » favoriteCuisine(P,C)

  In our future work, we would like to embellish this rule to accommodate personal choices as well - so a person's favourite cuisine will be a combination of the social influence on them and their personal tastes.
- "socialInfluenceOnCuisine(A, +B) = 1 ."

  We believe that there's a functional constraint on a influences on a person such that if a person gets more and more influenced towards a particular cuisine, then his influence towards other cuisine decreases by an equivalent amount
- "favoriteCuisine(A, +B) = 1 ."

  Similarly, a person's preferences of cuisines is bounded such that if he/she starts liking one cuisine more, then the likeliness towards other cuisines decreases an equivalent amount

We extended our model by adding following rules. The motivation was to capture a user's popularity and convincing ability when they influence culinary choices of their friends. We captured popularity of a user from Yelp attributes for a user such as useful votes, cool votes and funny votes.[KC14]

- **5**: (Friend(P1,P2) & favoriteCuisine(P1,C) & usefulUser(P1)) » socialInfluenceOnCuisine(P2,C)
- **5**: (Friend(P1,P2) & favoriteCuisine(P1,C) & coolUser(P1)) » socialInfluenceOnCuisine(P2,C)
- **5**: (Friend(P1,P2) & favoriteCuisine(P1,C) & funnyUser(P1)) » socialInfluenceOnCuisine(P2,C)

5 Evaluation

To evaluate the collective classification model, we created observation and truth partitions in the dataset. We hid 20% of the observed cuisine labels which were kept in the truth file. The labels were later compared to PSL results to calculate performance of the model. We evaluated the model based on accuracy of the overall model and precision, recall and F1 score for individual classes. We experimented by removing and adding various rules while evaluating the model. Column one indicated presence of the popular votes rules in the model. The second column indicates presence of functional constraints. We evaluated the model using various sizes of samples. The table 2 shows results of evaluation performed on sample of 600 users.

Our belief for why removing the popularity rules did better than keeping them on is as follows - since usefulUser, cooluser, and funnyUser are artifacts of Yelp, the friends network is not too influenced by them since they know the user from before and on a more personal level owing to the fact that they were friends on Facebook before Yelp. And so, they essentially are an overhead.

6 Future Work

There are a few things we would liked to have tried, but weren't able to. But, in the future, we would definitely like to:

- Instead of treating this as a classification problem, we would like to propose a ranking model wherein we rank the gold favourite cuisines of an individual by the frequency of their occurrences in the review text. So, if a person talks about his predilection towards cuisine X more than cuisine Y, X is ranked higher than Y in his cuisine preference order. So, we arrange the predicted labels in the descending order of their truth values and compare them to the gold labels reporting a ranking loss
- We would also like to vary the amount of labelled and unlabelled data in the collective classification task. This will also help us understand how to best selected the labelled entities for maximum propagation, and in essence help find a way to identify influential Yelp users.
- We would also like to implement some form of user similarity measure so that we could implement some user-similarity rules in our PSL model.

References

[CW12] Julius Cheng Calvin Wang, Mirela Spasova. Predicting and modeling activity in the evolving yelp network. 2012.

[HT16] Suzanne Higgs and Jason Thomas. Social influences on eating. *Current Opinion in Behavioral Sciences*, 9:1–6, 2016.

[KBB+12] Angelika Kimmig, Stephen Bach, Matthias Broecheler, Bert Huang, and Lise Getoor. A short introduction to probabilistic soft logic. In *Proceedings of the NIPS Workshop on Probabilistic Programming: Foundations and Applications*, pages 1–4, 2012.

[KC14] Johhny Winston Kevin Crain, Kevin Heh. An analysis of the "elite" users on yelp.com. 2014.

[LKJJ06] Bae-Hee Lee, Heung-Nam Kim, Jin-Guk Jung, and Geun-Sik Jo. Location-based service with context data for a restaurant recommendation. In *International Conference on Database and Expert Systems Applications*, pages 430–438. Springer, 2006.

[PPC08] Moon-Hee Park, Han-Saem Park, and Sung-Bae Cho. Restaurant recommendation for group of people in mobile environments using probabilistic multi-criteria decision making. In *Computer-Human Interaction*, pages 114–122. Springer, 2008.

[yel] Yelp dataset challenge dataset. `https://www.yelp.com/dataset_challenge/dataset`. Accessed: 2017-04-27.