

# Local Elites in Yelp

Brian Cheang, Brandon Kates

October 22, 2017

## 1 Research Context

High ratings and awards can drive large tourist crowds into local favorite restaurants, often causing restaurants to change (increase prices, new booking rules, impersonal service) to better accommodate the new customer base. As a result, the most popular and highest-rated restaurants may no longer be true local favorites but instead thrive on their popularity with tourists. By identifying local users and local experts to provide ratings reflective of local opinions, Yelp may become more popular with tourists looking to "travel like a local" and enjoy a more authentic experience.

## 2 Research Objective

The objective of the project is to build and combine two models (Local Expert Identifier / Topical Expert Identifier) for the purpose of identifying 'experts' among yelp users. The "Local Expert Identifier" is a Gaussian Mixture Model that identifies clusters in a given user's review locations to predict the user's most probable location. The "Topical Expert Identifier" is currently a supervised learning algorithm that combines different features about the users reviews in a certain category in order to determine if they are an expert in that category. The goal is to see if an unsupervised algorithm would be able to classify users into clusters of expert and non-expert without needing labels. The goal is to combine the models to find local experts in a specific category.

## 3 Research Road Map

### 3.1 Experiment Design

To build the "Local Expert Identifier", we plan to compile all provided review locations for each user in the dataset. Using each user's corresponding locations, we will fit a Gaussian Mixture Model to identify clusters (starting with 12 components). We will then iteratively reduce the number of components until the model has identified the key distinct clusters. Then, the user's most probable location will be determined using the mean of the most probable mixture component. To further improve the model, the model will check that 'local expert' users: 1. Visited the city over a longer span of time (Using review dates) 2. Has a majority of reviews in the city (No. of reviews in the city / Total no. of reviews)

To build the topical expert model, we first combine the business, user, and review tables in order to get a feature set for finding an expert in a category. Currently, we are using supervised learning, specifically, Random Forests, Gaussian Naive Bayes, and Decision Trees. The only issue is that these required labeled data in order to fit the model, so the goal in the future is to move towards unsupervised learning and see if 'clusters' of feature sets start to form. Currently, we classify elite users as the experts and everyone else as not being an expert.

This method requires building a different model for each category because the categories are so different and so it is unlikely that someone is an expert in all categories.

Currently, the feature sets we are using do not take into account the reviews themselves, and so the further direction is to use the length of the reviews as well as the 'relevance' or 'quality' in order to determine if someone is truly an expert.

### **3.2 Research Timeline**

10/16 - 10/22: Replicate Models in Jindal Thesis

10/23 - 10/29: Improve Models

10/30 - 11/12: Perform analysis on reviews using our 'Expert Identifying' models (phrase-level sentiment analysis)

11/13 - 11/19: Compile findings with relevant visualizations and begin writing final paper

11/20 - 11/26: Write final paper

11/27 - 12/2 - Submit final paper

### **3.3 Resources**

## **References**

- [1] Jindal, Tanvi. Finding Local Experts from Yelp Dataset. University of Illinois at Urbana-Champaign. Published 2015-04-27. <http://hdl.handle.net/2142/78499>