

Local Elites in Yelp

Brian Cheang, Brandon Kates

November 8, 2017

1 Abstract

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. Using information about review locations, review dates, and review counts, a model can be built to separate users into locals and 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 Introduction

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 Local Authority Model

3.1 Previous Research

Brief overview of GMM Local Expert model used in "Finding Local Experts From Yelp Dataset"

3.2 Proposed Improvements

- Filtering by time spans
- Filtering by review ratios

3.3 Evaluation of Model

4 Topical Authority Model

4.1 Previous Research

Brief overview of Topical Expert model used in "Finding Local Experts From Yelp Dataset"

4.2 Proposed Improvements

4.3 Evaluation of Model

5 Conclusion

5.1 Reflection

Explanation of why models were successful or unsuccessful in producing significant results. Include interesting discoveries made throughout the research.

5.2 Analysis of Yelp Dataset using combined model

If models are successful in identifying "expert" users, we will apply models (Local, Topical, Combined) to Yelp Dataset to analysis of Yelp reviews.

6 Resources