Phrase-Level Sentiment Analysis on Local Experts in the Yelp Dataset

Abstract

The Yelp dataset provides us with a limited amount of information from each user, but it would be interesting to infer more information about each user to get a better sense of what types of users exist on Yelp. Specifically, it would be interesting to take a look at where people are writing reviews, and try to 'local experts' in specific categories that could provide the best possible reviews for someone who is just in town visiting that day. We can take a look at both experts in specific categories, and experts in the neighborhood. It would also be interesting to look at the actual reviews that the users write, and to do sentiment analysis on local experts in order to come up with a summary for a business that gives potential customers a better feel for the business.

Introduction

In the first paper, the author describes a method for determining a user's location including the notion of topical authority and local authority. That is, whether or not the user has sufficient 'expertise' in a category, and whether or not the user has expertise in the local area.The paper only classifies users as local experts if they are classified as experts in both categories. The author uses a mapreduce algorithm to get a list of locations that each user reviewed, and then fits the data to a Gaussian Mixture Model (GMM) using Expectation Maximization (EM). This identifies location centers where a user's reviews fall. This is used to find local authority based on the distance a business falls from the location center.

The author uses a number of different features to categorize topical authority, concentrating around Yelp age, number of friends, average rating, variance of ratings, review length. It also uses similar features to identify user expertise in a category, and even delves into correcting for bias by comparing the distribution of a single user's reviews to the distribution of all user's reviews for the local business, the category, and the population. The author uses a different model for each category, and takes Yelp Elite Users as the ground truth for examples of experts.

The second paper looks at phrase-level sentiment analysis and using Collaborative Filtering (CF) in order to personalize recommendations to a specific user.

The methods in the two papers could be used to find a local expert that best matches another user.

Data/Design

Need to use review.json, user.json, business.json, and even potentially tip.json if we want to analyze more text. We want to use the Yelp dataset in order to classify users as local experts, and then look at the reviews from local experts to do some kind of sentiment analysis and determine the best and worst aspects of a business, and aggregate those into some kind of review. Going to use the Yelp dataset to experiment with the current models proposed in the papers, and also expand to generate other models to see how they perform.

Methods

1. Gaussian Mixture Model and Expectation Maximization
    a. Going to play around with the GMM and EM model presented in the first paper in order to get a sense of how it can be improved, play around with parameters.
2. Naive Bayes, Decision Trees, Random Forests
    a. These are the algorithms used in the first paper, going to try these as well as others such as decision trees, neural networks.
3. Collaborative Filtering based on Matrix Factorization Techniques

Significance

This work is useful for people that use Yelp in order to find the best local businesses, and it is useful for the owners of businesses on Yelp in order to improve their customer's experience. For example, in the case of a restaurant a local user could find local experts reviews of the restaurant and use those in order to determine whether or not they want to try to restaurant. Additionally, the user could find local experts that may share common food tastes with them, and that could influence how they view the local business.

Additionally, for the business itself, the reviews of local experts could shape changes that the restaurant might want to make in order to improve the negative parts that experts talk about and to enhance the positive parts. The local experts reviews could be aggregated and summarized for the local business to see these positive and negative aspects.

Finally, any efficient and effective algorithms used to characterize the yelp dataset could be applied to many other datasets.

References

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

Zang, Yonfeng. Incorporating Phrase-level Sentiment Analysis on Textual Reviews for Personalized Recommendation. State Key Laboratory of Intelligent Technology and Systems Department of Computer Science and Technology Tsinghua University, Beijing, 100084, China. Published February 2015. https://dl.acm.org/citation.cfm?id=2697033

Chee Hoon Ha. Yelp Recommendation System Using Advanced Collaborative Filtering. Stanford University. http://cs229.stanford.edu/proj2014/Chee%20Hoon%20Ha,%20Yelp%20Recommendation%20System%20Using%20Advanced%20Collaborative%20Filtering.pdf