Progress Report One

bjk224, bmc78

October 29, 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 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 Methods and Results

4 Individual Work

Brian

- Ran initial model (only GMM clustering) on subset of 1000 users to create sample data
- Added new variable "time_span" to users in sample data

- Began analysis to determine viable "cutoff" time for filtering tourists

Brandon

- Built a topical identifier model that initially used supervised learning techniques such as Naive Bayes, Random Forests, and Decision Trees, trained on elite users as 'experts'. Evaluation done by using a section of that data for training and for testing.
- Improved upon initial model by using K-means clustering with 2 clusters to try and identify differences in two sets of yelp users.

5 Further Studies

To evaluate the performance of the local expert model, we plan to create a baseline model that predicts a user's rating for a business given the following information: ("user_id", "business_id") Then the baseline model will be compared to a similar model that is trained on "user_id", "business_id", AND "is_local" (whether the user is a local from the area of the business as determined by our local expert identifier model).
In order to improve the topical expert model, we are exploring different feature sets that give more robust interpretations of users, and also working on evaluating the effectiveness of unsupervised learning models.

6 References

References

[1] Jindal, Tanvi. Finding Local Experts from Yelp Dataset. University of Illinois at Urbana-Champaign. Published 2015-04-27.

[2] 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.

[3] Chee Hoon Ha. Yelp Recommendation System Using Advanced Collaborative Filtering. Stanford University.