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. 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. To achieve this, we propose two different models: a local authority model and a topical authority model. The approach is evaluated using the Yelp dataset.

**Introduction**

The objective of the research 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 each 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 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.

The objective of the project is to build a model to classifying users as experts in locality and category. The trouble with the dataset is that there are no labels for user locality or experts among users. Therefore, we employ unsupervised learning techniques to try and find clusters of users that fit in one of the two categories. In this paper, we propose both local authority and topical authority models to identify these local experts. The "Local Authority Model” is a Gaussian Mixture Model that identifies clusters in each user's review locations to predict the user's most probable location. The “Topical Authority Model” employs k-means clustering on a set of users who have written reviews for a specific category to determine if the user is an expert in that category.