Travelling Like a Local: Extracting local opinions from Yelp reviews

Brian Cheang

September 24, 2017

Abstract

Often times, high ratings and awards drive new tourist crowds into local favorite restaurants. However, often times the restaurants change (increase prices, new booking rules, impersonal service) to better accommodate the new customer base. As a result, these restaurants are no longer truly local favorites but instead become tourist traps that drive away locals and instead thrive on high ratings of tourists. In this paper, we create an algorithm for calculating a restaurant’s local rating and tourist rating to provide both customers and business owners with more insight.

1 Research Context

Currently, local restaurants are rated by a single averaged star rating (0 – 5) from Yelp reviews of all types of users. As a result, the highest rated restaurants in the area may be catered to tourists and are not truly “local favorites”. Although many researchers have toyed with the idea of generating personalized Yelp reviews or even star ratings for specific aspects (food, service,…), there is not much being done to distinguish between true local experiences and tourist traps.

2 Data / Design

I plan to use data from review user, and business to combine information of users’ ratings of businesses and restaurant location and date of the users’ visits. The user data from the Yelp dataset does not include the user’s location so I am assuming the user is a local if the user has several reviews of restaurants in a city over a long span of time. The design of this research is to devise an algorithm to determine local ratings and tourist ratings of restaurants and apply this algorithm to gain insights on restaurants.

3 Methods

To determine where a user is from, I plan to use a Gaussian Mixture Model that takes coordinates and attribute corresponding locations to by determining the center of their review clusters. These locations are then filtered by the range of dates of the reviews (probably visiting if the reviews only span one week). With information about user locations, I plan to separate each restaurant’s reviews into reviews done by locals and reviews done by tourists/visitors.

To analyze the difference between local and tourist opinions, I plan to use time-series analysis to analyze the average stars given by locals/tourists/all over time.

4 Significance

By providing insights to local tastes, Yelp can become more appealing to tourists looking to travel like a local instead of visiting the same few trendy spots dominated by other tourists.

5 References

<https://www.ideals.illinois.edu/bitstream/handle/2142/78499/JINDAL-THESIS-2015.pdf?sequence=1&isAllowed=y>

http://delivery.acm.org/10.1145/2510000/2507163/p165-mcauley.pdf?ip=128.84.126.36&id=2507163&acc=ACTIVE%20SERVICE&key=7777116298C9657D%2EB493315FA1EC298D%2E4D4702B0C3E38B35%2E4D4702B0C3E38B35&CFID=814160865&CFTOKEN=74270786&\_\_acm\_\_=1506632189\_482ff160fe008a5abe640c766bfc0da4