Optimal Location For a New Business

Harika Sabella, Mohammad Adil, Sugeerth Murugesan

May 12, 2014

1 Introduction

Yelp is an online urban guide and business review site with over 120 million unique viewers each month and 53 million local reviews. We will try to answer an important social question by using this datset; "What is the location where we should place a new business to generate optimal profit? How much does location influence the prosperity of a business?"

1.1 Approach

The Yelp dataset contains nodes which represent businesses, users, reviews, check-ins, and tips. We will visualize the Yelp dataset through a structure closely related to bipartite graphs–we have sources which represent different communities (based on the zip code we determine if a business falls into a certain community) and each community contains business nodes (nodes which represent businesses based on latitude and longitude) and each business node will have multiple users reviewing it. Each edge from the business to the user will either be a positive edge or a negative edge where a positive edge indicates the positive reviews that the business gets and the negative edge indicates the negative reviews that the business gets–we will determine whether or not a review is positive or not through semantic analysis. Then, given a new business node that is to be placed within a certain community,we begin the effort to find the optimal location for this node by first, finding the source node that corresponds to the zip code of the community. Then, using the geographical layout of the businesses in that community, we determine the optimal location to place the new business node. We can make this determination based on a lot of factors–certain business analysts suggest placing a new business closest to a business that has the most positive reviews, so that customers will be more likely to visit the new business because it is close by to an already popular business. Other business analysts suggest that a new business should be placed in a location that is central to all the other businesses of the same genre. We could use the center sites algorithm to determine where this central location is–the centers would contain all the businesses belonging to the same genre in the community and the site would be the new business node.

2 Resources

We essentially get our data from contest "Yelp dataset challenge". The dataset contains useful information about user reviews, ratings, check ins and tips of businesses located in Phoenix, Arizona. We plan make a visualization of the Yelp data in iGraph, and derive statistical models using R.

3 Literature review

The paper Inferring Future Business Attention talks about a method called sentiment analysis which is run on the user provided reviews to pick a few of the best reviews (determined based on the reviews which have the most comments) out of all of the given reviews and make a conclusion about the future business success. The paper "Mining Opinion Features in Customer Reviews" is about how to determine whether or not a review is positive. The paper Hidden Factors and Hidden Topics: Understanding Rating Dimensions with Review Text provides an algorithm that determines how a user would react to a new product based on the products that the users interests which are determined by the investments the user makes on already existing products.

References

[1] Bryan Hood, Victor Hwang and Jennifer King, *Inferring Future Business Attention*

[2] Julian McAuley,Jure Leskovec *Hidden Factors and Hidden Topics: Understanding Rating Dimensions with Review Text*

[3] Ken Slonneger Chapter 5, Minqing Hu and Bing Liu *Mining Opinion Features in Customer Reviews*