Fast Food Retail Store Location Selection by Predictive Model

Laurence Lin

2020

1. Introduction

1.1 Business Problem Background

In the business field, selecting an appropriate location for new opening store is an important task when entering an unexplored market. Opening a new store near the city center or the outskirt of town may lead to huge difference on business profit, which is highly related to business's success. In order to find a suitable location for a new store, retailers and business analyst should made lots of research, including surveying local market, number of competitors, the cost to invest a new store and so on. These business analysis cost a lot of time and money, and will result in a huge loss if the new store location is decided improperly. Nowadays, since big data and machine learning technique has been more developed and maturely applied in various business field, I decide to bring up a method helps select the best retail store location by the provided geographic data and predictive model. This may help business stakeholders and retailers save time and cost to get better understanding to local market, and help with better location of new retail store.

1.2 Problem Definition

Elements that may affect a store's income including consumer behavior, consumer demographics and competitiveness of other stores. Foursquare API provides location based data such as competitors, number of other venues and other geographic data within local area. Though the check-in data provided by Foursquare API is not rich in certain location, we apply the geographic data from Foursquare API, along with retail store popularities and ratings from Google Place API as target data. Our aim is to apply location based data from Foursquare API to rank the location candidates by predicting popularity and ratings of certain location.

1.3 Interest

Retailers and investors may want to find a method to intelligently select an optimal location for a new retail store, reducing the risk of financial loss.

Analyst from other business field may also be interested if they want to start a new business of investment in an new unexplored area.

2. Data Description and Collection

2.1 Data Description

Assuming a fast food retailer wanted to open an new store in a city, and investigates the consumer potentials and competitiveness. The retailer analyzes the relationship between geographic data and profits of existing fast food store, and build a mathematical model with it. In this project, I selected Mcdonald's restaurant as the retail store to analyze. I explored the geographic data near the restaurant in New York, and build a predictive model to estimate the popularities and ratings for the new store location candidates.

Locational based service has provided informations for data scientist to analyze tasks around a certain area. Foursquare API, for example, enables us to scrape geographic data around certain venue and human activities. For retail store location selection task, business analyst considers factors such as demography, incoming flow, cuisines habit, competitiveness and lots of elements. I mine Foursquare geographic data that may be informative about the retail quality of a geographic location.

2.2 Data Collection

The Foursquare geographic data is collected around each retail store, within the area of radius 200 meters. The whole range of analysis covers an circle area with radius 10 kilometers, centroid locate in geographical center of New York(Manhattan area 40°46'26.2"N 73°58'49.9"W). For all the retail stores located in this range, I collected the geographic data around each store. The geographic data in the vicinity of retail store is assumed to be informative for the store's popularity or ratings. For example, a retail store opened near railway station implies more potential consumers, while a retail store open near lots of competitors may reduce it's consumer. As for each store's popularity and ratings, I applied Google Place API to extract the ratings and number of comments for each store as indications of the location quality.

The Foursquare geographic data contains informations of density, heterogeneity, competitiveness, and area popularity. Density represents the

number of nearby venues. Heterogeneity calculated the neighbor entropy, indicates the diversity of category venues in the surroundings. Competitiveness is measured by number of same category venues nearby the retail store, in this case including number of fast food restaurants and other restaurants. Area popularity is calculated by number of residential venues around the store, which may imply number of residents. Given the analyze area set $\{p \in P,\ dist(p,\ l) < r\}$ where p stands for each nearby venue around location l within radius r, the geographic features is measured as follows:

Density: Number of all venues p around store location l.
$density = |\{p \in P,\ dist(p,\ l) < r\}|$

Heterogeneity: Define number of neighbor venue with type γ as $N_\gamma(l,\ r)$, total venues in the area $N(l,\ r)$, measure entropy of all place types in the area.

$neighbor\ engropy = -\sum_r \frac{N_\gamma(l,\ r)}{N(l,\ r)} * log(\frac{N_\gamma(l,\ r)}{N(l,\ r)})$, for all types γ in the area.

Competitiveness: Compute number of same type venues around the retail store, here including type *fast food restaurant* and *other restaurants*.
$Competitiveness = -\frac{N_{\gamma,l}(l,\ r)}{N(l,\ r)}$, $N_{\gamma,\ l}(l,\ r)$ is number of venues of type γ same as location l.

Area popoularity: Assume number of residentail places around location l indicates the popularity.
$Area\ popularity = N_{res}(l,\ r)$, $N_{res}(l,\ r)$ is number of residential venues around location l.

The retail quality of an geographic area is measured by popularity and ratings, we aimed to build a model that could predict high ranking locations based on the retail quality. First we assess the performance of each geographic feature based on ability to predict highest retail quality spots, observe the performance of each feature, then we combine these features to observe if the model improves.