The Battle of Neighborhoods

Manas Ranjan Pradhan

Solution

- **Business Hubs**

  Locations need to discovered which are Business hubs in Mumbai city.

  Using Foursquare location data, points of interests for entire Mumbai can be extracted.

  We shall not specify any category while downloading data for points of interest. After downloading this data we can create clusters based on the coordinates of the points.

- **Clustering**

  There are various clustering techniques available. Density based algorithm comes to mind. This shall divide the points based on density. Instead we can specify a certain number of clusters by our self. It shall help in dividing the points in more evenly.

Decision for Location

Locations needs to decided considering following factors

- **Near by cluster**

  The business hub or cluster which is nearer to home may be preferred for starting the business.

- **Cluster not having a type of business**

  If there are clusters where there are less or no presence of business in my preferred category, those can be suitable.

- **Commonly found together**

  As an extended approach, we can analyze the clusters and find with which type of business the preferred category is found together. There might be some unknown benefits of having those businesses nearby. We can check if those are present our preferred cluster.

Data Description

- From Foursquare location data, gps coordinates for shall be downloaded using API for entire region.
- Category and name column denote the type of business for each of the points of interest. These two columns can be used together.

  Eg.

  Category “New Café coffee shop” represents a business type of “coffee shop”

  Name “Great delicious bakery” represents a business type of bakkery.

- We my need to filter out irrelevant data like ‘building’, ‘space’ etc.
- Data extraction may be a heavier process and can not be repeated frequently. We shall save the extracted and processed data in flat files in different steps during our analysis.