The Battle of Neighborhoods

Gustavo C. Bicalho

gustavocbicalho@gmail.com

February 10, 2019

Overview

Introduction

Motivation:

- Moving to a new city can be difficult.
  - No information of the venues in each neighborhood.
  - What if we had an application that selects a neighborhood according to our interests?

Solution:

- A Content based recommendation system that helps the user choosing a neighborhood.

Data

- Toronto neighborhood and boroughs dataset - As an example city, using a dataset with all boroughs, postal code and name of each neighborhood.
- Geospatial data for Toronto - Contains the latitude and longitude for each neighborhood in the city of Toronto.
- Foursquare API - Allows to obtain the data on what venues are located at each neighborhood. In this data, we will be able to obtain the latitude and longitude for each venue, as well as the category and name for them.
- User data - a random user will be generated, with a random number of preferences(from 1 to 10) and select randomly from the list of categories available in the city (obtained through the Foursquare API) that number of categories.

Methodology

- Beautiful Soup - Extract the Toronto postal code, boroughs and neighborhoods information from an HTML file.

	Postcode	Borough	Neighborhood
0	M1B	Scarborough	Rouge, Malvern
1	M1C	Scarborough	Highland Creek, Rouge Hill, Port Union
2	M1E	Scarborough	Guildwood, Morningside, West Hill
3	M1G	Scarborough	Woburn
4	M1H	Scarborough	Cedarbrae

Figure: Figure showing the table with the postal code, borough and the grouped neighborhoods.

Methodology

- Geospatial data from Toronto - latitude and longitude for the neighborhoods in Toronto

	Postcode	Borough	Neighborhood	Latitude	Longitude
0	M1B	Scarborough	Rouge, Malvern	43.806686	-79.194353
1	M1C	Scarborough	Highland Creek, Rouge Hill, Port Union	43.784535	-79.160497
2	M1E	Scarborough	Guildwood, Morningside, West Hill	43.763573	-79.188711
3	M1G	Scarborough	Woburn	43.770992	-79.216917
4	M1H	Scarborough	Cedarbrae	43.773136	-79.239476

Figure: Figure showing the table with the postal code, borough and the grouped neighborhoods and the longitude and latitude for them.

Methodology

- Foursquare API - Get the information on which venues are available at each neighborhood
  - Apply OneHot encode to obtain a table with the neighborhood as the index and the categories as the columns, where the value in each column is how much that categorie is relevant in the neighborhood.

Methodology

- Random User - Generate a random user that will be used to test the system
  - Select a random number of categories for the user.
  - Create a table with all categories as the columns and one row, where the values are 1 if the user has interest in that category or 0 otherwise.

Methodology

- Recommendation System - Multiply the user profile with the table that has the neighborhood and the weight of each category.
  - Result in a matrix with the score of each neighborhood.
  - The higher the score, the better that neighborhood fit the user interest

Results and Discussion

- In this prototype, we'll generate a random user for the tests:

```
['College Stadium', 'Airport Food Court', 'Nightclub', 'Optical Shop']
```

Figure: Categories random selected for the user.

Results and Discussion

	Postcode	Borough	Neighborhood	Latitude	Longitude	Score
0	M1N	Scarborough	Birch Cliff, Cliffside West	43.692657	-79.264848	0.250000
1	M5V	Downtown Toronto	CN Tower, Bathurst Quay, Island airport, Harbo...	43.628947	-79.394420	0.071429
2	M6G	Downtown Toronto	Christie	43.669542	-79.422564	0.062500
3	M5S	Downtown Toronto	Harbord, University of Toronto	43.662696	-79.400049	0.028571
4	M7A	Queen's Park	Queen's Park	43.662301	-79.389494	0.022222

Figure: Score table obtained for the random user.

Results and Discussion

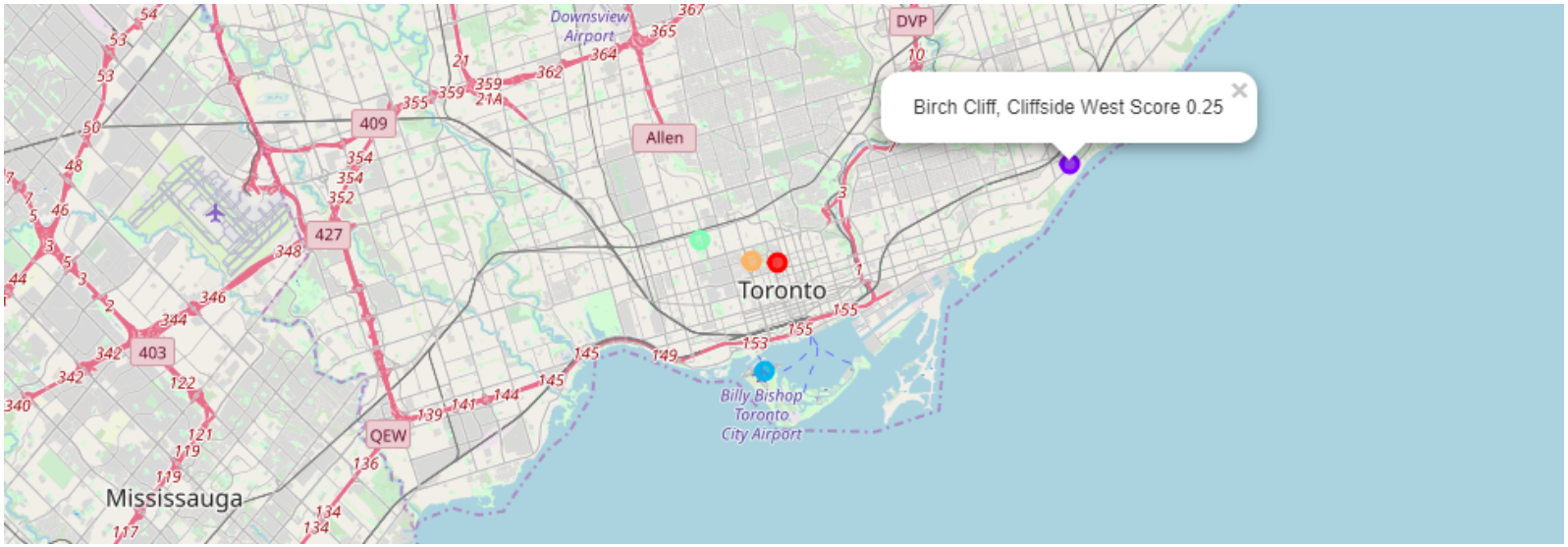

Figure: Map with the 5 best neighborhoods for our user.

Results and Discussion

- The best neighborhood based on his interests would be "Birch Cliff, Cliffside West".
- The difference between the first and the other results obtained are large.
  - Unusual categories such as 'Airport Food Court' and 'College Stadium'
  - "Birch Cliff, Cliffside West" is the neighborhood where College Stadium has the highest weight.

Conclusion

- Simple content based recommendation system that still need improvements.
  - If a region has few venues assigned to it, the weight of single venues will be huge resulting and a probably best fit.
  - It's possible to consider only neighborhoods with a certain number of venues found.
  - Also, we are not considering other possible interests:
    - How close the place is to the users work.
    - GDP of the neighborhood
- This is only a prototype, the idea is that the application can be used for different cities and allow the user to enter the categories they prefer.

The End