Neighborhood pairings: how to move to a new city

David Robinson

May 14 2020

1. Introduction

When moving to a new city, it takes some time to get to know the place and get oriented. Particularly in major metropolitan cities, which neighborhood you live in matters quite a lot - each one has its own atmosphere, like a miniature city. How, when moving somewhere for the first time, could you pick a neighborhood? As a solution, I'm proposing a tool to map neighborhoods in one city to the most similar neighborhoods in another. To prove the concept, I will use the Toronto and New York.

2. Data

2.1 Data sources

For this task, Foursquare data will be essential. The specific data will be all venues in both cities (unless I limit this for speed and proof of concept.) For example, a sample row of Foursquare data that I use would need at minimum the latitude, longitude, name, and category of a venue. I will also need GeoJSON files of Toronto and New York for visualization.

2.2 Use of Data

Using this data, I'll determine, for each neighborhood in New York, which is the most similar neighborhood in Sydney - and vice-versa. Instead of using a clustering algorithm for this, since I want a one-to-one mapping, I'll simply calculate the euclidean distance between all the neighborhoods (once they've been properly encoded.) The neighborhood in the new city with the smallest distance from the neighborhood in your current city will be the best choice!

3. Methodology

The first steps to this project involved loading and cleaning the data. GeoJSON was loaded, and combined with the Foursquare data into a dataframe for each city. To simplify the problem, I narrowed the scope to look at only one Borough for each city. To handle categorical data, I one-hot encoded the venue categories provided by Foursquare.

In order to compare Euclidean distances between neighborhoods, I first needed neighborhoods from the two cities to be of the same dimension. Because of its size, Manhattan had higher dimension (more venue categories.) To handle this I dropped all venue types that weren’t present in both cities.

Finally I was able to calculate the Euclidean distance between neighborhoods using scipy. This gave a matrix of distances between all neighborhoods. The only remaining step to find the closest neighboring city is to take a row or column min.

Results

The key result is the distance matrix representing distance between all pairs of neighborhoods. With this, it is trivial to create a tool that would help users pick a neighborhood in their new city, and the original objective is met.

Discussion

The best way to test the accuracy of these claims would probably be to consult someone with local knowledge of both cities. However, it is safe to say that the results could be improved with additional data on the neighborhoods. Venues alone is probably not enough, since crime, price and even weather are all key when picking the right neighborhood.

In terms of the pairing algorithm, using Euclidean distance was exceptionally simplistic. However, for the problem at hand, it may be the best option, as it simply found the nearest neighbor neighborhood.

Conclusion

In conclusion, the tool seems to be feasible. Someone moving from Toronto to New York or New York to Toronto could quickly put their favorite neighborhood into the tool and get a new neighborhood suggestion right back out. This shows the proof of concept, and next steps would be to add more data and get the tool ready to handle major cities anywhere.