Finding similar neighborhood

Based on the works done in –

"Where is the Soho of Rome? Measures and Algorithms for Finding Similar Neighborhoods in Cities. “

-G´eraud Le Falher, Aristides Gionis, Michael Mathioudakis

Karan Jeet Singh (A20327513)

Kanwal Preet Singh (A20327416)

Problem

- Wanted to answer questions like “Where is the soho of rome?”.
- Look for similar neighborhoods across different cities.
  - Studies done on one neighborhood could potentially be applied to its counterpart.
  - Find accommodation in the kind of neighborhood you are familiar with.
  - Find comparable issues, and their potential solutions.

Approach

- We solve the problem in three stages –
  - First we learn distance between venues and evaluate them in information retrieval fashion.
  - Using the historical ground truth data of neighborhoods we find out parameters of the distance measure
  - Then similar neighborhood are discovered using the Earth Mover’s Distance Euclidean.

Data

- We collected data from foursquare via twitter.
- Combining the historical data and the data we gathered. [1]
- Total of 30K checkins for Chicago, and 20K for Barcelona.
- Via twitter–
  - Foursquare checkin ID, venue id, foursquare user id, loc, city, time, tweet id, tweeter user id, twitter message body.
- Via Foursquare API –
  - Name, loc, category, checkin counts, users count, tip counts, price, rating, mayor, tags, shortUrl, canonicalUrl, likes, likers, city, closed, hours.

[1] Le Falher, Géraud; Gionis, Aristides; Mathioudakis, Michael (2015): Foursquare & Flickr activities in 20 cities. fig**share**. http://dx.doi.org/10.6084/m9.figshare.1584973Retrieved 00:00, Nov 30, 2015 (GMT)

Results

- We were able to identify similar neighborhood across two different cities
- EMD-EUCL is the best performing measure for finding the similar neighborhood
- EMD-LMNN performs only slightly worse.
- The results are not very close to ground truth.
- Non-circular shaped ground truth data causes low overlap and hence low relevance score.

Conclusions

- Using EMD-EUCL we can find similar neighborhoods across different cities.
- Improvements-
  - Use and analyze image data (Flickr, Instagram, etc.)
  - Match several neighborhoods at the same time and extend that to match cities.
  - Include the data for air quality, weather and transportation.