Finding Similar Neighborhood

Characterizing neighborhood based on the geotagged foursquare, twitter data and location references, then doing sentiment analysis on that data and also finding features of that neighborhood. Then using the data to find similar neighborhood for different cities.

"Where is the Soho of Rome? Measures and Algorithms for Finding Similar Neighborhoods in Cities. -Géraud Le Falher, Aristides Gionis, Michael Mathioudakis"

Kanwal Preet Singh – A20327416
Karan Jeet Singh – A20327513

Problem

- Moving between cities is always a difficult task because you don’t know the people or the place.
- It is easier to be in a neighborhood which is similar to your previous neighborhood.
- We are tackling this problem by helping users to find a similar neighborhood across cities which could make their stay more pleasant and enjoyable. Also it will help us learn about the different environment a city has to offer.

Approach

- We implement an approach developed by Géraud Le Falher, Aristides Gionis, and Michael Mathioudakis.
- Using geo enabled data from social-media platforms, we represent each venue with a feature vector, that accurately describes the characteristics and the overall activity of the venue. We then devise similarity measures between venues, as well as between neighborhoods, i.e., sets of venues that are geographically close to each other.
- We address these two problems from a metric-learning point of view (Bellet, Habrard, and Sebban 2013). The earth-mover's distance (EMD) (Rubner, Tomasi, and Guibas 1998). EMD is known to be a robust measure—however, it is also expensive to compute. Motivated by this observation, we address the issue of computational efficiency. In particular, given a neighborhood R in one city, we ask how to find the k most similar neighborhood to R in another city (or a set of other cities) under EMD, and without performing brute force computation. We implement a pruning strategy that yields speed improvement with minimal loss in accuracy.
- Our study and our algorithms are based on extensive experimental evaluation in European and US cities, using activity logs gathered from Foursquare via Twitter.

Data

- Foursquare is a popular location-based social network that, as of 2015, claims more than 50 million users.
- It enables users to share their current location with friends, rate and review venues they visit, and read reviews of other users. Foursquare users share location information by generating “check-ins” using a dedicated mobile application.
- Each check-in is associated with a web page that contains information about the user, the venue, and other details of the visit. Each venue is also associated with a public web page that contains information about the venue—notably its category such as Food or Nightlife Spot—and aggregates information from user check-ins. According to Foursquare’s privacy policy, check-ins are private information. However, sometimes users opt to share their check-ins via Twitter, a popular microblogging platform.

Timeline

- Phase 0 (18th October 2015 – till project ends)
  - Collect twitter data of Chicago, New York, Paris, London, and Barcelona.
- Phase 1 (18th October 2015 – 30th October 2015)
  - Create algorithm and technique to extract feature vectors.
- Phase 2 (31st October 2015 – 13th November 2015)
  - Create algorithm and technique to define venues by assigning them feature vectors.
  - Create algorithm and technique to classify neighborhoods and then cities based on venues.
- Phase 3 (14th November 2015 – till project ends)
  - Testing and evaluating results on various neighborhoods of different cities.

Credits

- Source - "Where is the Soho of Rome? Measures and Algorithms for Finding Similar Neighborhoods in Cities. -Géraud Le Falher, Aristides Gionis, Michael Mathioudakis"