What would you do next?
Modelling Activity Transitions using the Foursquare API

Elson Serrao   ELSON.SERRAO@GMAIL.COM

1. Introduction

Have you ever wondered, where do people go after watching a game at the stadium? Or if you just ate lunch what would you do next? If you were at the game and your team won you would probably end up at a bar or pub celebrating your team's victory. And after lunch, you would probably want to get some dessert. Are these just guesses, or are they backed by some statistical data?

The main goal of this project is to answer such questions, leveraging the data obtained from location sharing services like Foursquare. Each individual task that a person does could be considered as an activity. For example, watching a game at the stadium, having drinks at a bar, eating lunch, having dessert are all activities. The transition from watching a game at the stadium to having drinks at a bar or eating lunch to having dessert is considered to be an activity transition. In this project, we will try to model the activity transitions and answer the primary question: "If a person is doing some known activity currently, what is the person most likely to do next?"

Modelling activity transitions could be considered as a sub-task of modelling patterns of human mobility which are significant for traffic forecasting, urban planning, as well as epidemiological models of disease spread [Cheng et al. (2011)]. As such one of our target audiences would be governments and municipalities. Understanding human mobility also allows developers to enhance recommender systems [Noulas et al. (2011)] providing targeted recommendations to users.

2. Related Work

As compared to what research is already present in this field, this is a very small scale project. If the topic has piqued your interest, I would suggest you to go through the papers listed below. However, definitely have a look at Livehoods if the papers seem boring. They have a very good interactive web interface that help understanding several aspects of a dynamic city.

Related work in this field usually utilize the checkin footprint of users to model human mobility. However, getting checkin information using the Foursquare API is a premium call which caused several limitations. Also unless the checkin is made public, the API terms and conditions states that you cannot store the data more than a couple of hours. For this reason, I had to mostly use only the regular calls.

2.1 Data Sources

As a use case, in this project we will try to model activity transitions in the city of San Francisco. We would first need geospatial data to get the boundaries of San Francisco. It can easily be obtained from DataSF. We get the geojson files for zip codes and San Francisco neighbourhoods.

We will then use Google Maps Geocoding API to get locations within the zip codes and neighbourhoods. Using these locations as seed points we will explore the venues of San Francisco using the Foursquare API. A venue is identified by the venue_id. We parse the result obtained from the explore endpoint of the Foursquare API to store the venue name, address, category, city, state country, postal code, cross_street, latitude and longitude. A venue is capable of having several categories. However, every venue has one category marked to be it's primary category. We extract only the primary category for the venue.

For each venue, we use the next venues endpoint from the API to get a list of next venues. This endpoint gives us the top 5 most probable venues to be visited next. An activity transition could be considered as a change in venue i.e. when a person moves from one venue to the next. And the activity the person preforms at the venue is the primary category of the venue. We can then correspondingly map the venue transitions to activity transitions using the category to which the venue belongs to.

We consider the primary category of the venue as the activity. The list of all the categories available in Foursquare are also obtained. Each category will be identified by its unique id. A category can have several sub-categories. Along with the id, we retrieve the name of the category, the icon that represents it and the parent category if it has any.

For examples of the data please have a look at the notebook Data Description

References

Zhiyuan Cheng, James Caverlee, Kyumin Lee, and Daniel Sui. Exploring millions of footprints in location sharing services, 2011. URL `https://www.aaai.org/ocs/index.php/ICWSM/ICWSM11/paper/view/2783`.

Anastasios Noulas, Salvatore Scellato, Cecilia Mascolo, and Massimiliano Pontil. An empirical study of geographic user activity patterns in foursquare, 2011. URL `https://www.aaai.org/ocs/index.php/ICWSM/ICWSM11/paper/view/2831`.