Predicting Real-time Population Density of Singapore based on Social Sensors

Ashraf Mohammed Abdul - A0123411B	Astha Jain - A0123438J
Harshul Gandhi - A0123448H	Jabir Shabbir - A0123480N

Introduction

Census data is a valuable source of information. It provides states with a unique insight into the various factors affecting their populace and can play a key role in policy and governing decisions. However, census data is static ie a snapshot at a moment in time and performing a census, especially for large states is such an expensive and time consuming affair that the data gathered is outdated by the time the findings are published.

We propose to supplement census data with information gathered in real time from social sensors to predict the real time population density for a given geographic region. State activities such as disaster management and planning, state surveillance, infrastructure planning may benefit from such an analysis.

Methodology

We plan to extract data from social sensors and event listing websites. The extracted data will then be analyzed with respect to user demographics, social pressures and weather information to calculate the number of people attending these events for a particular area. These raw numbers will then be extrapolated to the population density data from census to help predict the real-time population density for Singapore. Following are the detailed steps involved:

1. Crawl different event listing websites and gather list of events happening in a particular area. Target websites for this step are: yoursingapore, eventfinda, craiglist.
2. Extract all the check ins and posts from social networking websites such as facebook and twitter for events gathered in step 1. We will use twitter and facebook APIs to perform this step.
3. Search for each of these event's page on these social networking websites and get the list of users listed as 'attending'. We are going to use facebook APIs for this.
4. We will then try to assert if a particular user is really going to attend an event based on multiple social and geographical factors such as: Interest, Peer Pressure, weather, distance, time of day, financial, popularity, health.
5. Output of above 4 steps would be a calculation of the number of people that are going to attend a particular event in a given location.
6. We will then try to relate the calculated number from the step above with the census population in a particular area.
7. Our final target is to generate a real-time prediction as a heat map of population density in Singapore.

Predicting Real-time Population Density of Singapore based on Social Sensors

Ashraf Mohammed Abdul - A0123411B	Astha Jain - A0123438J
Harshul Gandhi - A0123448H	Jabir Shabbir - A0123480N

Novelty

Most accurate way of performing such an analysis is through real time cellular network data. However this approach has not widely been implemented because of such data is not shared at state level due to privacy concerns and lack of consensus among network operators and state entities. In the absence of this data, publicly available social sensor information might be able to provide a crude approximation.

Tools

1. Python
2. MySQL
3. R
4. HTML5
5. Javascript

Data Sources

1. Facebook API
2. Twitter API
3. Foursquare API
4. Eventfinda - http://www.eventfinda.sg/whatson/events/singapore
5. YourSingapore - http://www.yoursingapore.com/festivals-events-singapore.html
6. craiglist - http://singapore.craigslist.com.sg/search/eve
7. stclassifieds- http://www.stclassifieds.sg/events-and-concerts/entertainment-and-events/list/c72
8. whatshappening - http://www.whatshappening.sg/events/
9. population area wise-http://data.gov.sg/
10. past weather data-http://aws.amazon.com/datasets/2759