M-ATTRACT: ASSESSING THE ATTRACTIVENESS OF PLACES BY USING MOVING OBJECTS TRAJECTORY DATA

André Salvaro Furtado\textsuperscript{12}, Renato Fileto\textsuperscript{2}, Chiara Renso\textsuperscript{3}

\textsuperscript{1}DG/UDESC, \textsuperscript{2}PPGCC/UFSC, \textsuperscript{3}KDD-Lab, CNR – Pisa
Overview

- 1. Introduction
- 2. Related Work
- 3. Definitions
- 4. Proposed Measures
- 5. Discussion and Analysis
- 6. Conclusion and Future Work
1. Introduction

- **Attractiveness** quantifies how much something is able to attract the attention and influence the decisions of individuals [Uchino et. al. 2005]

- In the real world it is possible to identify places that are more attractive to the individuals who move around them
  - These places tend to influence the space around them
    - Ex.: a shopping mall
1. Motivation

- Attractiveness quantification and understanding are useful for several areas:
  - Urban Planning, Transport, Marketing, Market, Analysis, Migration Analysis, Tourism

- Assessing attractiveness of places based in real trajectory data can improve the understanding of the geographic space configuration and dynamics
  - Trajectory data can be automatically collected and more detailed than survey data
2. Related Work

- **Attractiveness** is studied since the early 30’s
  - Attraction between cities
    - *Gravitational Attractiveness Model* [Reilly 1931]
    - *Central Place Theory* [Cristaller 1933]
  - Other contemporany approaches usually:
    - Adapt these theories
      - Using static or estimated values (e.g., population, purchasing power, area, distance)
    - Based on survey data
      - Subjectivity
2. Related Work

- Widespread use of mobile devices (e.g., GPS) enable the collection of large data volumes
  - Real data about the movement around places can be used to estimate their attractiveness

  **Trajectories**
  - Each **raw trajectory** $\tau$ is a time ordered sequence of points:
    $\tau = \{\mu_1,\ldots, \mu_n\}$
  - And each point $\mu_i$ is a triple:
    - $\mu_i = (x_i,y_i,t_{si})$ where:
      - $x_i, y_i$ are spatial coordinates
      - $t_{si}$ is a time instant
2. Related Work

- **Episode** is a maximal segment of a trajectory that comply to a given predicate \cite{Mountain and Raper 2001}
  - Extracted by processing raw trajectories integrated with geographic data
    - E.g., passed near somewhere, is stopped at a place...

- **Semantic trajectory**
  - \cite{Spaccapietra et al. 2008}
    - A set of relevant places visited by the moving object
  - \cite{Parent 2012}
    - A trajectory that has been enhanced with annotations and/or one or several complementary segmentations

![Diagram of a trajectory with places marked: Airport 9pm, Shopping A, Shopping B, Hotel]
2. Related Work

- The use of trajectories to evaluate the attractiveness still is in its first steps

- [Giannotti et al. 2007] [Giannotti et al. 2011]
  - Regions of interest built from a grid-based partition of the space into rectangular cells
    - Aggregates adjacent cells whose measures of trajectories concentration around them are similar
2. Related Work

- [Wei et al. 2010]

  - Framework for pattern-aware trajectories mining
    - Uses the density-based algorithm introduced in [Giannotti et al. 2007] to extract regions that are passed by at least a certain number of trajectories
    - Attractiveness of trajectories from the attractiveness of the regions
      - A trajectory is considered more attractive if it visits more regions with high attractiveness
2. Related Work

- Works based on the analysis of taxi trajectories
  - [Yue et al. 2009]
    - Clusters that group spatial temporal similar pick-up and drop-off points of taxi trajectories
    - Attractiveness of clusters based on the time-dependent flows between clusters
  - [Yue et al. 2011]
    - Assess the attractiveness of shopping centers using:
      - Data about them (i) and about trajectories (ii)
        - (i) e.g., gross leasable area, number of nearby shopping malls, available parking space
        - (ii) e.g., number of taxis within the shopping’s area of influence in different time periods
## 3. Proposed Method

<table>
<thead>
<tr>
<th></th>
<th>Categories of Places</th>
<th>Trajectories of Individuals</th>
<th>Scale of Analysis</th>
<th>Trajectory Episodes</th>
<th>Support of Episodes Parameteres</th>
</tr>
</thead>
<tbody>
<tr>
<td>Reilly [1931]</td>
<td>Predefined Regions</td>
<td>X</td>
<td>Single</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>Cristaler [1933]</td>
<td>Predefined Regions</td>
<td>X</td>
<td>Single</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>Gianotti [2007][2011]</td>
<td>Created Regions</td>
<td>Any</td>
<td>Single</td>
<td>Passing</td>
<td>No</td>
</tr>
<tr>
<td>Wei [2010]</td>
<td>Created Regions</td>
<td>Any</td>
<td>Single</td>
<td>From/To</td>
<td>No</td>
</tr>
<tr>
<td>Yue [2009]</td>
<td>Created Regions</td>
<td>Taxi</td>
<td>Single</td>
<td>Drop in / Drop off</td>
<td>No</td>
</tr>
<tr>
<td>Yue [2011]</td>
<td>Shopping Mall Only</td>
<td>Taxi</td>
<td>Single</td>
<td>Drop in / Drop off</td>
<td>No</td>
</tr>
<tr>
<td>M-Attract</td>
<td>Predefined and/or Created Regions</td>
<td>Any</td>
<td>Multiple</td>
<td>Stops and Passing</td>
<td>Yes</td>
</tr>
</tbody>
</table>
3. Definitions: Region, Subregion and Place

- **Spatial scope**: a three level hierarchy
- **Automatically or manually defined**
  - **Region of Interest \( r \)**
    - The totality of the analyzed space
  - **Subregion of Interest \( s \)**
    - Non-overlapping portions of the region of interest that are relevant for the attractiveness of places analysis
  - **Place of Interest \( \rho \)**
    - Portions of the subregions of interest considered in the analysis
    - The atomic unity in the analysis

- **E.g. 1**, Florianópolis \( r \), Centro \( s \), Beiramar Shopping \( \rho \)
- **E.g. 2**, Beiramar Shopping \( r \), 1st floor \( s \), Burger King \( \rho \)
3. Definitions: stopAt, passBy and passIn Episodes

- Three different kinds of episodes with different parameters:
  - stopAt(τ, ρ, ξ, δ)
  - passBy(τ, ρ, ξ)
  - passIn(τ, ρ)
4. Proposed Measures

- Four different measures to capture different aspects of the attractiveness phenomena
  - All the measures are based on the sum of the number of episodes previously described
    - We ensure that all measures give a value in the interval [0, 1] if the denominator is greater than 1
      - If the denominator is 0 by convention the given value is 0
4. Examples Data Set

- **Datasets**
  - Region extracted from GADM
    - Milan City
  - Subregions extracted from Wikimapia
    - 40 neighborhoods inside Milan City
  - Places extracted from OSM database
    - 16,044 places inside the selected subregions
  - Trajectories of **private cars** inside the Milan City
    - Over 10k trajectories
      - Collected between 5th and 7th July, 2007
  - Episodes detected using a generalization of SMoT Method [Alvares 2007]
    - Buffer = **30m**
    - Minimum staying = **120s**
4. Distribution of Episodes

- Higher density of episodes in central areas
- StopAt Episode
- PassBy Episodes
4.1. Measure 1 (for places): Absolute Stopping Capacity (ASC)

- Proportion of the numbers of stopAt and passBy episodes calculated for a given place \( \rho \)

- Aspect
  - Capture the attraction of a place over the movement that passes around it

- Consideration
  - Places with a low number of stopAt episodes located at low movement areas will also have high values

\[
ASC(\rho, \Gamma) = \frac{\sum_{\tau \in \Gamma} \text{Count}(\text{stopAt}(\tau, \rho, \xi_\rho, \delta_\rho))}{\sum_{\tau \in \Gamma} \text{Count}(\text{passBy}(\tau, \rho, \xi_\rho))}
\]
4.1. Distribution of ASC values

Central Area
Low values

Peripheral Area
High values
4.2. Measure 2 (for places): Relative Stopping Capacity (RSC)

- Proportion of the number of stopAt of a given place $\rho$ and the total number of stops in all places spatially contained in the same subregion $s$

- **Captured Aspect**
  - Place’s popularity weighted by the popularity of a given subregion $s$

- **Consideration**
  - Don’t capture the movement difference between different subregions

$$RSC(\rho, s, \Gamma, \Phi) = \frac{\sum_{\tau \in \Gamma} \text{Count}(\text{stopAt}(\tau, \rho, \xi_\rho, \delta_\rho))}{\sum_{\tau \in \Gamma, \rho' \in \Phi} \text{Count}(\text{stopAt}(\tau, \rho', \xi_{\rho'}, \delta_{\rho'}))}$$
4.2. Distribution of RSC values
4.3. Measure 3 (for subregions):  
Relative Passing Stopping (RPS)

- Product of the proportion of episodes `passIn` and `stopAt` in a given subregion `s` over the region `r`

- **Captured Aspect and Consideration**
  - The movement and the attraction of this movement inside a subregion compared to the whole analyzed space
  - Allow the differentiation between subregions to improve the analysis of the places contained in them

\[
RPS(s, r, \Gamma, \Phi) = \frac{\sum_{\tau \in \Gamma} \text{Count}(\text{passIn}(\tau, s))}{\sum_{\tau \in \Gamma} \text{Count}(\text{passIn}(\tau, r))} \times \frac{\sum_{\tau \in \Gamma, \rho' \in \Phi} \text{Count}(\text{stopAt}(\tau, \rho', \xi_{\rho'}, \delta_{\rho'}))}{\sum_{\tau \in \Gamma, \rho'' \in \Phi} \text{Count}(\text{stopAt}(\tau, \rho'', \xi_{\rho''}, \delta_{\rho''}))}
\]
4.3. Distribution of RPS values

- Darker colors mean higher values

- Conurbation Area
- Central Area
- Rural Area
4.4. Measure 4 (for places): Strict Attractiveness (SA)

- Product of the previously seen measures (ASC, RSC and RPS)
- **Captured Aspect and Consideration**
  - Uses all the other measures to find the most attractive places
    - High attraction of the movement around
    - High attraction considering all the movement attracted by places in the same subregion
    - Located at an attractive and movemented subregion

\[
SA(\rho, s, r, \Gamma, \Phi) = ASC(\rho, \Gamma) \times RSC(\rho, s, \Gamma, \Phi) \times RPS(s, r, \Gamma, \Phi)
\]
5. Discussion and Analysis

- Is possible to distinguish groups of attractive places in several regions
- The most attractive places are well distributed around the city being subregion's big attractors
5. Discussion and Analysis

- **Commercial x Residential zones**
  - Different patterns are found
  - The comparison of ASC values and number of stopAt episodes in different zones is noticeable
    - Residential zones tend to have high ASC values and a low number of stopAt episodes
5. Discussion and Analysis

**Left – Top 10 places by Absolute Stopping Capacity**

<table>
<thead>
<tr>
<th>Place Name</th>
<th>StopAt</th>
<th>PassBy</th>
<th>ASC</th>
</tr>
</thead>
<tbody>
<tr>
<td>Apartments (Via P. Fiuggi, 19)</td>
<td>5</td>
<td>5</td>
<td>1</td>
</tr>
<tr>
<td>Starhotels Tourist</td>
<td>6</td>
<td>6</td>
<td>1</td>
</tr>
<tr>
<td>Apartments (Viale dell’Aviazione, 62)</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Apartments (Via Donna Prassed, 2)</td>
<td>6</td>
<td>6</td>
<td>1</td>
</tr>
<tr>
<td>House (Via Privata Faiti, 9)</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Apartments (Via Val Maira)</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Apartments (Via Luigi Bertelli)</td>
<td>6</td>
<td>6</td>
<td>1</td>
</tr>
<tr>
<td>Asilo Nido</td>
<td>2</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>Apartments (Via San Mirocle)</td>
<td>6</td>
<td>6</td>
<td>1</td>
</tr>
<tr>
<td>House (Via Gaetano Crespi)</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
</tbody>
</table>

**Right – Top 10 places by Strict Attractiveness**

<table>
<thead>
<tr>
<th>Place Name</th>
<th>StopAt</th>
<th>PassBy</th>
<th>SA</th>
</tr>
</thead>
<tbody>
<tr>
<td>Metropoli</td>
<td>154</td>
<td>177</td>
<td>0.00198</td>
</tr>
<tr>
<td>Bicocca Village</td>
<td>70</td>
<td>83</td>
<td>0.00098</td>
</tr>
<tr>
<td>Esselunga di Via Ripamonti</td>
<td>80</td>
<td>109</td>
<td>0.00097</td>
</tr>
<tr>
<td>Esselunga di Via Rubattino</td>
<td>38</td>
<td>81</td>
<td>0.00082</td>
</tr>
<tr>
<td>Esselunga – Missaglia</td>
<td>40</td>
<td>43</td>
<td>0.00062</td>
</tr>
<tr>
<td>Mediaworld</td>
<td>24</td>
<td>48</td>
<td>0.00055</td>
</tr>
<tr>
<td>Mango Italia</td>
<td>43</td>
<td>95</td>
<td>0.00041</td>
</tr>
<tr>
<td>Galleria Manzoni</td>
<td>45</td>
<td>109</td>
<td>0.00039</td>
</tr>
<tr>
<td>Esselunga di Via Novara</td>
<td>34</td>
<td>51</td>
<td>0.00038</td>
</tr>
<tr>
<td>Milano Centrale</td>
<td>58</td>
<td>261</td>
<td>0.00036</td>
</tr>
</tbody>
</table>
6. Conclusion and Future Work

- An initial step in a new approach to assess the attractiveness of places based on trajectory data

- **Main advantages:**
  - Flexibility
    - Scale, Places, Individuals, Episodes, Detection Parameters
  - A variety of measures to capture different aspects of the attractiveness
  - Real movement data of individuals

- **Future Work:**
  - New efficient algorithms to detect episodes on large datasets
  - Semantic enrichment of geographical and trajectory datasets
  - Temporal Aspects
References


References


