Impact of accessibility indices on secondary activity type and location type prediction using random forest classifiers

Author(s): Ordóñez Medina, Sergio A.

Publication Date: 2016-09-14

Permanent Link: https://doi.org/10.3929/ethz-b-000120637

Rights / License: In Copyright - Non-Commercial Use Permitted
Impact of accessibility indices on secondary activity type and location type prediction using random forest classifiers

Sergio ORDONEZ
Future Cities Laboratory (FCL),
Singapore-ETH Centre for Global Environmental Sustainability (SEC), Singapore
Email: ordonez@ivt.baug.ethz.ch

How accessible are locations in space and time, is a property widely used and studied in transportation science. Accessibility is a measurement of this property, a number which quantifies how easy is to reach a certain location. If a location is not accessible new transport supply (roads, public transport services, etc.) should be implemented. Accessibility indices are also used for location choice models (e.g. Srour et al. 2003), health or education coverage (e.g. Yang et al. 2006), and population distribution (e.g. Song 1996), among others. Litman 2007 shows that, regarding Transport Planning, accessibility is affected by mobility (physical movement), quality and affordability of transport options, transport system connectivity, mobility substitutes, and land use patterns.

However accessibility is not a physical quantity, and it can be defined and estimated in many ways. Although there is a well-established form to calculate accessibility in transportation, shown in Equation 1, many others ways have been proposed. Handy and Niemeier 1996, Bath et al. 2001, and Geurs and van Wee 2004, among others, review differences in accessibility indices for transportation and land use. The original definition was designed to be applied to zones, but it can easily be adapted for individual buildings. The cost function is commonly associated to distance decay functions or distance distributions obtained from data (e.g. Martinez and Viegas 2013).

\[ A_i = \sum_j O_j \cdot f(C_{ij}) \]

*Equation 1: Accessibility of the place \( i \) depends on \( O_j \) which is the attractiveness or number of opportunities of the place \( j \), and a function of \( C_{ij} \) which is the cost to reach the place \( j \) from the place \( i \).*

This paper studies the influence of the main locations accessibility to secondary activity locations, as a predictor of secondary activity type and location-type. Several accessibility definitions are compared to find the best predictors for each classification model. According to the Singaporean Household Interview Travel Survey (HITS) performed in 2012, and a database of more than 250 thousand locations, 8 location types were related to the 4 most common secondary activities in the survey. Figure 1 presents this relation. It’s important to point out that several location-types can be assigned to one location (in Singapore each house or building has a different postal code), and each type is weighted according to the number of opportunities within the specific building. Thus, a big shopping mall may have a high weight for the type “shop_high” and “eat_high”, but a low weight for “sport”.

In order to compare the impact of different accessibility definitions, the first step consists of calculating 6 accessibility indices of the main locations of the HITS respondents (home, school,
work place), by place-type. In other words, for each main location of a person who performs a secondary activity in the survey, 48 accessibility indices were calculated, corresponding to 6 definitions by 8 place-types.

On one hand, these 6 accessibility definitions were tested independently as predictors of the observed secondary activities. Given an accessibility definition, the corresponding indices of a respondent’s main locations were used to predict which secondary activity he/she will perform. In other words, each time a trip registered in HITS has a purpose included in the set of secondary activities (see Figure 1), accessibility measurements of the origin and final destination are used as predictors, and the observed activity is used as an outcome, to train a random forest classification model. Random forests have been used to recognize activity purposes (see Liu 2013). Thus, each accessibility definition feeds a different random forest model, and the differences in the accuracy of the classification models is used to compare the impact of the accessibility definitions on the secondary activity prediction.

On the other hand, a similar approach is used to compare the predictability strength of the accessibility indices on location-type. Given one accessibility definition one classification model per secondary activity type is trained (i.e. four random forest models are trained per accessibility definition). To train each of these models it is assumed that when a person selects a certain secondary activity he/she chooses a place type restricted to the relation presented in Figure 1. This means that, given one accessibility definition, a trip observed in HITS with a certain secondary activity purpose, is just used to train one corresponding random forest classification model. Given a secondary activity, the 6 location-type classifiers (one for each accessibility definition) can be compared using their accuracy, in the same manner of the activity classifiers.

Figure 3 shows relative errors of the secondary activity classifiers using the 6 different accessibility definitions. The lowest errors not always are achieved with the same accessibility definition. For certain activities local accessibility could be a better predictor than more general definitions. The paper with discuss and present in detail more results.
Figure 2: Six accessibility indices of the main locations of Singapore to high-demand shopping places. a.) Opportunities by travel time, b.) Travel time by opportunities, c.) 50 more “Opportunity by travel time” places, d.) Opportunities times Cost function, e.) 50 local “Opportunities times Cost function” f.) 5 local “Opportunities times Cost function”

<table>
<thead>
<tr>
<th>Accessibility</th>
<th>Act</th>
<th>shop</th>
<th>eat</th>
<th>errands</th>
<th>rec</th>
</tr>
</thead>
<tbody>
<tr>
<td>a.) Opportunities by travel time</td>
<td>19.4%</td>
<td>10.4%</td>
<td>4.5%</td>
<td>5.2%</td>
<td></td>
</tr>
<tr>
<td>b.) Travel time by opportunities</td>
<td>9.9%</td>
<td>12.7%</td>
<td>5.9%</td>
<td>7.5%</td>
<td></td>
</tr>
<tr>
<td>c.) 50 more “Opportunity by travel time”</td>
<td>12.2%</td>
<td>12.0%</td>
<td>4.8%</td>
<td>6.5%</td>
<td></td>
</tr>
<tr>
<td>d.) Opportunities times Cost function</td>
<td>13.3%</td>
<td>12.0%</td>
<td>5.7%</td>
<td>7.0%</td>
<td></td>
</tr>
<tr>
<td>e.) 50 local “Cost function”</td>
<td>12.1%</td>
<td>11.9%</td>
<td>4.6%</td>
<td>6.0%</td>
<td></td>
</tr>
<tr>
<td>f.) 5 local “Cost function”</td>
<td>11.3%</td>
<td>10.6%</td>
<td>4.6%</td>
<td>5.2%</td>
<td></td>
</tr>
</tbody>
</table>

Best: b.) a.) a.) f.)

Table 1. Random forest secondary activity classifiers error by accessibility definition.
References


How accessible are locations in space and time, is a property widely used and studied in transportation science. Accessibility is a measurement of this property, a number which quantifies how easy is to reach a certain location. However accessibility is not a physical quantity, and it can be defined and estimated in many ways. This paper studies the influence of the main locations accessibility to secondary activity locations, as a predictor of secondary activity type and location-type. Several accessibility definitions are compared to find the best predictors for each classification model. According to the Singaporean Household Interview Travel Survey (HITS) performed in 2012, and a database of more than 250 thousand locations, 8 location types were related to the 4 most common secondary activities in the survey. In order to compare the impact of different accessibility definitions, 48 accessibility indices were calculated for each main location of a person in the survey, corresponding to 6 accessibility definitions by 8 place-types. For each secondary activity trip in the survey, accessibility measurements of the origin and final destination are used as predictors, and the observed activity is used as an outcome, to train a random forest classification model. Thus, each accessibility definition feeds a different random forest model, and the differences in the accuracy of the classification models is used to compare the impact of the accessibility definitions on the secondary activity prediction. A similar approach is used to compare the predictability strength of the accessibility indices on location-type. Given one accessibility definition one location-type classification model per secondary activity type is trained. Given a secondary activity, 6 location-type classifiers (one for each accessibility definition) can be compared using their accuracy, in the same manner of the activity classifiers. Results show that the lowest errors not always are achieved with the same accessibility definition. For certain activities local accessibilities could be a better predictor than more general definitions.