Commuting time for every employed: combining traffic sensors and many other data sources for population statistics

Pasi PIELA
Statistics Finland, Finland

As part of accessibility studies in today’s GIS (geographic information science and systems) research, Statistics Finland has implemented a national road and street database for applications based on population statistics.

Computationally most complex are network optimisations at micro data level. In 2014, commuting distances were upgraded to commuting drive time and drive time optimized distances by utilising standard models and traffic sensors. Specifically, data accumulated by traffic sensors are to complement speed estimation during rush hour traffic.

The actual pair wise computations to estimate shortest paths have been made for all 2.1 million employed persons that have coordinate pairs. These statistics are being updated annually at micro level as part of the Social Statistics Data Warehouse.

Accessibility differs from area to area. Thus, a modern and detailed GIS based urban-rural classification is applied in the analysis.

Keywords: network analysis, road database, population statistics, big data, sensor data
1. Data Warehouse

Digiroad is a national database managed by the Finnish Transport Agency, which contains accurate data on the location of all roads and streets in Finland, covering a total of 483,000 km (Finnish Transport Agency, 2013). It has been successfully implemented earlier for accessibility applications based on population statistics of Statistics Finland, such as elementary school accessibility, emergency accessibility and commuting distances (Piela, P. 2013). This paper will concentrate on the upgrade of commuting distances to commuting time itself.

The extensive Social Statistics Data Warehouse of Statistics Finland offers dwelling coordinates and work place coordinates for citizens along with a variety of demographic features (Nieminen, J. & Myrskylä, P., 2010). The work-place coordinate coverage was 91.2 per cent of all inhabitants in 2011 and 92.4 per cent in 2008. In future, it is expected that fewer people will have a fixed singular coordinate pair due to the changing nature of the “work-place” (e.g. manpower leasing).

In these analyses the detailed Urban-rural classification by the Finnish Environment Institute SYKE is also applied at micro level.

2. Traffic detection data

The main challenge is to find out average travel time for each road element (here the smallest unit in the road network) in any part of the whole country. Generally, speed limit based models are considered naïve especially during the rush hours. A local example on a sophisticated approach for travel time modelling has been given by the research group on Spatial Patterns on Accessibility of the University of Helsinki. It has done extensive research and produced free software that calculates the travel time from any selected point to another within the Helsinki Metropolitan Area. It takes into account the time of day, traffic lights and intersections, even the time that parking takes (Salonen M. & Toivonen T., 2013 and MetropAccess, 2013).

The Finnish Transport Agency also offers traffic sensor data for its customers and other specific user groups - open data services (Digitraffic.fi) being available as well. However, there are only 437 stations, vehicle detection loops, at the moment, as shown in Figure 1. Nevertheless, they are considered to represent well the main roads of the Finnish road network. Actually, there tend to be more of these stations in busier areas as well (see Fig.
2). They give information for speed, direction, length and class of a passing vehicle. Here, the obvious target is the speed of passenger cars excluding heavy vehicles.

**Automatic traffic measurement stations and main roads of Finland**

**Automatic traffic measurement stations in the metropolitan area of Helsinki**

*Figures 1 (upper) and 2. Traffic measurement stations in Finland and in the greater capital area along with high ways and trunk roads of Finland.*
3. Drive-time model

Traffic sensor data were selected for a four-week winter period between Finnish seasonal holidays: weekdays (Monday-Friday) from 13 January to 7 February 2014. The selected time interval is generally the busiest morning hour 7:00-8:00 based on annual statistics of these stations. Both directions are handled separately.

The main route network is divided into zones for generalizing (spreading) actual point observations equally along the same road within its main intersections and limiting the maximum distance from the point in question to less than 30 kilometers - such wide generalizations being possible only in sparsely populated areas. This method may obviously cause plenty of generalization errors and even bias. However, the actual generalized speed estimate (based on real observations) is controlled by comparing it to the real speed limit of a corresponding road element - each direction separately. The lowest speed is selected. Here, wintertime speed limits are used since Finland has a seasonal speed limit system.

The speed of all the other roads is estimated by using a specific road functional classification with 14 classes and strict computation speeds for each class varying between 20 and 95 km/h. The idea is to estimate the average speed so that it is lower than the actual speed limit and compensates traffic lights, intersections and other barriers slowing the traffic on streets within urban areas. Naturally, the route optimization algorithm itself uses hierarchical routing moving a target, “a car”, away from a slower street as soon as possible.

4. Pairwise computing of commuting times and distances

A straightforward approach for using Digiroad is given here: pairwise computing of commuting times for the approximately 2.1 million employed Finns that have coordinate pairs. The data required for such calculations are two years behind the actual target year (Piela, P., 2013).

The actual data for computations can be reduced by simple editing using linear distances. However, in this case, all the distances with a linear distance of more than 0.3 kilometers were included.

The ESRI ArcGIS software is used, of which Network Analyst Route Solver is called with a Python code. The method used is a common Dijkstra’s algorithm with hierarchical routing (ESRI, 2005), which makes calculation faster and is seen as a natural option here. The impedance attribute is the time and the hierarchy is now a six-class functional classification attribute (Class I main road, Class II main road, Regional road etc.).

The distance as a result is thus a drive time optimized shortest path to a work place being quite independent of the actual accuracy of the speed estimation method given the quality of Digiroad itself.
5. Commuting time results

The median ratio of a route distance and a corresponding point-to-point linear distance between a home and a main workplace is 1.35, the mean of ratios is 1.44, and the standard deviation is 0.42 meters (excluding the shortest routes of under 300 meters). Naturally, the ratio changes depending on the length class and geographical area.

As Table 1 indicates, the distributions of the commuting distance and time are skewed, the skewness of the commuting time being 3.82. Thus, the basic descriptive statistics shown are based on very robust measures: quartiles. Specifically, the dispersion measure is given as a quartile coefficient of dispersion (Bonett, D. G., 2006): the ratio of the interquartile range and the interquartile sum:

$$QCD = \frac{Q_3 - Q_1}{Q_3 + Q_1}$$

This is one robust and simple measure of variation that is used to make comparisons between sets of data.

*Table 1. Distance statistics in meters: median, mean of 0-200 km distances, 25th percentile, 75th percentile and quartile coefficient of dispersion (Q3 - Q1) / (Q3 + Q1).*

<table>
<thead>
<tr>
<th>Type</th>
<th>Median</th>
<th>Mean</th>
<th>Q1</th>
<th>Q3</th>
<th>QCD</th>
</tr>
</thead>
<tbody>
<tr>
<td>Linear</td>
<td>6.10</td>
<td>13.43</td>
<td>2.08</td>
<td>15.40</td>
<td>0.76</td>
</tr>
<tr>
<td>Route</td>
<td>8.91</td>
<td>17.04</td>
<td>3.13</td>
<td>20.40</td>
<td>0.73</td>
</tr>
<tr>
<td>Time</td>
<td>11.72</td>
<td>16.83</td>
<td>5.67</td>
<td>21.13</td>
<td>0.58</td>
</tr>
</tbody>
</table>

Figure 2a shows the area differences in commuting distances and Figure 3a in commuting time, respectively. Many remote areas tend to have shorter distances and times, on average, but the deviation is very high, as seen in Figures 2b and 3b. The metropolitan area of Helsinki, Helsinki Sub-region at LAU 1 level, indicated by a black square, which covers almost 30 per cent of the total employed population of Finland, has longer commuting distances than the average, not to mention the commuting time. The median distance for this metropolitan region is 11.37 kilometers; the median commuting time is 14 minutes 23 seconds.
6. Comparative results

SYKE’s new, detailed Urban-rural classification is utilized at micro level as seen in Figures 4a and 4b. It divides the country into seven classes: Inner urban, Outer urban and Peri-urban area, Local centre in a rural area, Rural area close to urban area, Rural heartland area, and Sparsely populated rural area.

It is clear as in Figures 4a and 4b that the commuting time depends on the urban-rural class in a way that for example populations in the inner urban area around the whole country tend to be similar from the commuting point of view. Both distances and time increase highly on neighboring rural areas. Rural areas in general have both very long and very short commuting times.

Similar differences occur when analyzing drive speed in these models. As an example, those who have their commuting distances between 10 and 15 kilometers are driving 47.7 kilometers an hour, on average, in the capital sub-region. However, those populations in the same region whose home coordinates are in the Rural heartland have 57.60 km/h as their average speed, still remaining behind of higher speeds of more northern and sparsely populated areas of Finland.

6.1. Time use survey

The only official statistics related to commuting time are produced along with the Time use survey (OSF 2014). Participating respondents keep a diary of their time use over a two-day period in units of 10 minutes. The last survey was conducted between 2009 and 2010. The definition of the commuting time concept is different here and based on respondent’s own experience: e.g. returning commuting period after leaving a workplace ends to the first stop over, e.g. grocery store.

Those data involving at least some minutes of commuting time (10 min., 20 min., 30 min. etc.) have their one-way lower quartile estimate in 10 minutes, the median in 20 minutes and the upper quartile in 30 minutes, the average value being 24.0 minutes (without extreme observations).

The current commuting time and distance data as presented in this paper are based on the employment information of 2011. Replacing shortest trips by a walking speed of 70 meters a minute (distances less than 1500 meters) and selecting thus observations with more than five minutes of travel time give the corresponding statistics as follows: 9.5 minutes (Q1), 14.7 minutes (Md), 22.7 minutes (Q3), the average being 22.5 minutes (without extreme observations).
7. Further research and conclusions

As part of the accessibility studies at Statistics Finland commuting distance and its upgrade to commuting time in the national route and street database are annually updated part of the Social Statistics Data Warehouse. Commuting time is a challenging problem that requires models of rush hour traffic. The traffic detection data support this challenge. The next update of the commuting distance and time data will be for the 2012 employment data in the end of 2014.

Further research in this field should include more the quality of life aspects along with comparisons to the Time use survey.

References


Figures 3a and 3b. The median and the quartile coefficient of dispersion of commuting distances optimized by drive time along the route network for populations of the sub-regions of Finland, at LAU 1 level, in 2011. Helsinki is marked with a black square also indicating also the surrounding Helsinki Sub-region. The classification is based on the Jenks optimization method.
Figures 4a and 4b. The median and the quartile coefficient of dispersion of commuting times along the route network for populations of the sub-regions of Finland, at LAU 1 level, in 2011. The classification is based on the Jenks optimization method.
Figures 5a and 5b. The median of commuting times populations in Inner-urban areas (left) and in Rural areas close to urban areas (right) by the sub-regions. The classification is based on Figure 2.