Analyzing a drivers stops and purpose from GPS data

Author:

Christian Juul Würtz, M.Sc.
Centre for Traffic and Transport, Technical University of Denmark

Abstract

The AKTA project was carried out in the period 2001 to 2003, examining the effects of introducing road pricing in the greater Copenhagen (Denmark) area. Data was collected as GPS points with 1 second intervals and mapmatched to a highly accurate map.

In the process of analyzing the change in the driver behavior, it was essential to distinguish between different kinds of stops. In other words it was important to differentiate between involuntary stops (e.g. stop because of traffic congestion) and deliberate stops (e.g. a parent setting of their children at a school).

A VB program that made extensive use of the buffer and clustering engine was developed to solve this problem. Based on a detailed land use map, time of day, time extent and location of other stops it was able to define the stop category and in part the trip purpose.

Introduction

The AKTA project was carried out in the period 2001 to 2003, examining the effects of introducing road pricing in the greater Copenhagen (Denmark) area[1]. All data was collected by an on board unit (OBU) as GPS points with 1 second intervals. A mapmatching was performed on a highly accurate map for later analysis

A part of this project was to analyze the change in driver behavior. This paper describes how to recognize stops that were not identified by OBU in the data collection phase and how the purpose of the trips was estimated.

Problems to solve

The dataset holds information from 300.000 trips, therefore an automated process was the only viable solution. It was not requested for the participants to keep a record of their trip destinations, it was therefore desired to estimate the travel purpose subsequently and to ensure that the trips were split according to what was defined as a stop.

The goal of this automated procedure was to accomplish the following:

1. To find stops that had not been recognized as stops in the data collection phase.
2. To estimate the purpose of the trips

The OBU settings for defining a stop was set to 180 seconds, in other words if the car was at a complete stop for more than 3 minutes the trip was split. This approach identifies most stops correctly. But in some cases e.g. where kids are set of at school, the vehicle is not stopped for more time than what is often the case in dense traffic situations. In other cases the OBU had as a result of technical problems not split the trips according to the 3 minute interval. To identify some of these missing trips and hereby correcting the data was an objective for the automated process.
At this point with the corrected data set, it was possible to analyze the stops with the objective of estimating the trip purpose.

Solution

First step before deciding the approach was to specify the data available.
The raw data in the form of GPS points at 1 second intervals was the appropriate data to analyze. This data is later mapmatchedand hereby simplified[2]. This has the outcome that the data is not suitable for localizing missing trips that are to be split or estimating the trip purpose. To determine the starting point of the trip, the last registered GPS point from the previous trip was accepted as the starting point of the current trip. This because the OBU and its incorporated GPS usually needs some time to start up, and therefore the first GPS point registered will in general be inaccurate. Along with the coordinates the raw data also includes a time stamp which permits the program to calculate the duration of the stop.
Highly accurate maps with attributes including road types, together with a land use map covering the area were also available.

The automated process was built by using VB in ArcGIS 8.1. Based on the available data it was necessary to define which parameter settings could identify the different kind of stops. The following characteristics were expected from the stop types

Characteristics used to localize a missing stop

- All stops enduring more than 60 seconds are considered as a potential stops.
- Will generally occur an smaller roads
- Are often repeated in the same area

Characteristics used to estimate trip purposes

- Home – The most frequent visited stop. Stop time is often above 12 hours and typically at night and on all weekdays. It is expected to be close to a residential area
- Work – stops between 5 and 12 hours. Usually not in weekends. Is probably not in a residential area.
- Other – Depending on the time ,day, time span ,location according to the landuse map and number of trips to the same area other purposes could be estimated

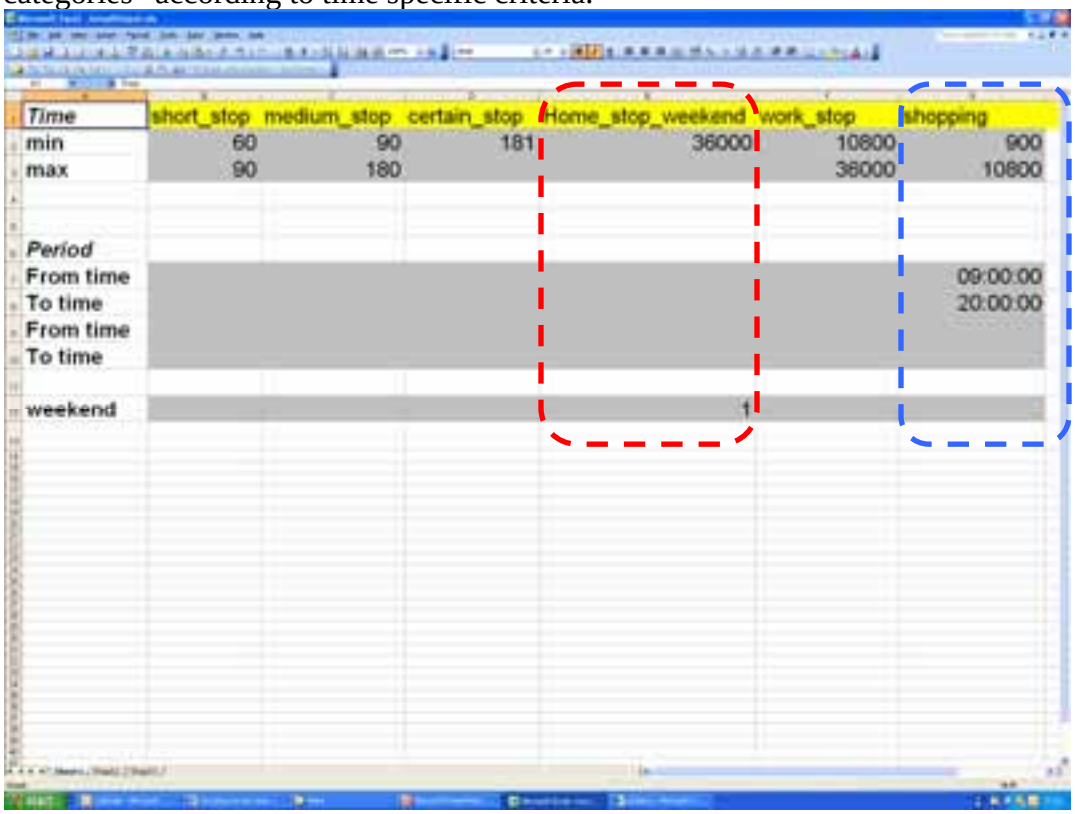
Program

The automated process was built using VB in ArcGIS 8.1. The user interface was made up of an excel spread sheet, to allow the user to create any number and type of definitions.

It was found that the best way to analyze the trips for the above characteristics was to divide the work into 2 phases.

Phase 1: Giving the stop at time dependent "category"

The first spreadsheet enables the user to provide each stop with any number of "categories" according to time specific criteria.



Time	short_stop	medium_stop	certain_stop	Home_stop_weekend	work_stop	shopping
min	60	90	181	36000	10800	900
max	90	180			38000	10800
Period						
From time						09:00:00
To time						20:00:00
From time						
To time						
weekend				1		

Figure 1 Spreadsheet user interface defining the stop category

Any number and form of "categories" can be created just by adding a new column to the spreadsheet. A stop can fall within any category or any number of categories.
With the settings set as in Figure 1, a stop longer than 36000 seconds (10 hours) with occurrence in the weekend will fall into the "Home_stop_weekend" (red dashed line). This is a way of identifying the residence stop. On the other hand if it's duration is between 900 and 10800 seconds (15 minutes to 3 Hours) and in the usual opening hours 9:00 to 20:00, it falls inside the "shopping" category (blue dashed line).
In this way the stops have been put into a "category" and are ready for phase 2

Figure 2 illustrates 4 stops. 3 of the stops are around 25 minutes and would be in the "shopping" category, whereas the remaining is not since it is only 6 minutes long.

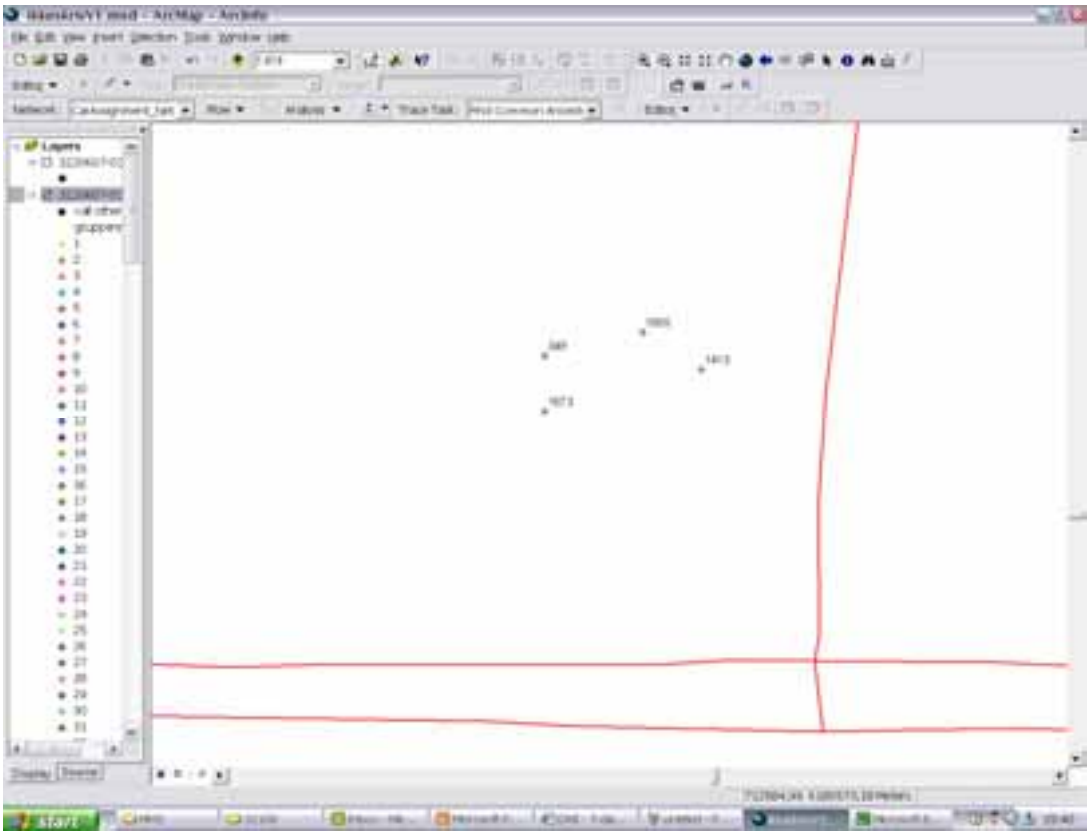


Figure 2 Four stops that are to be "categorized" by the program

Phase 2: Analyzing the stops according to their location

The second phase involves looking at the stops location in respect to each other and to the surrounding areas.
When a driver reaches a certain destination the coordinates registered by the OBU will vary due to many reasons.
GPS reception is always associated with some uncertainties especially under difficult reception conditions. Shopping in areas with huge parking lots or being resident in an area where parking is a problem as a result of the few parking spaces available, will often result in stops that although having the same purpose can be scattered over a vast area. This requires the program to utilize the clustering engine to make a connection between the stops and their destination. By calling the clustering engine the 4 stops illustrated in Figure 3 are predicted as having the same destination.

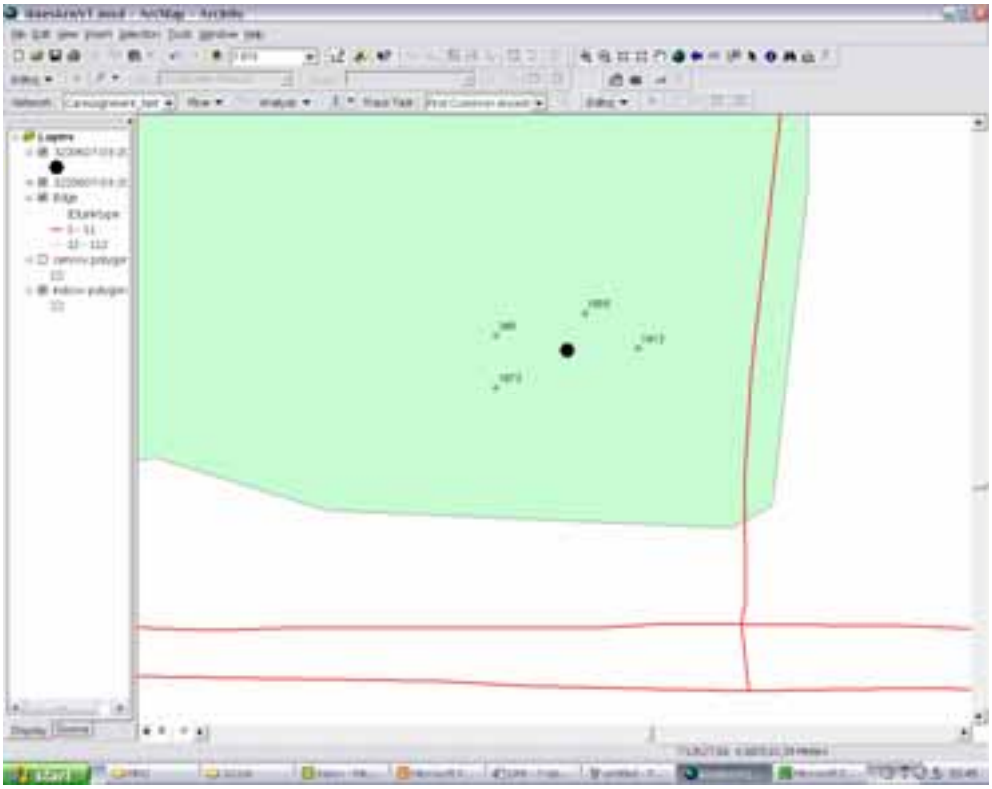


Figure 3 The 4 stop points are perceived as one destination (Black dot)

At this point the program perceives the data from the stops as one destination (Black point), containing all information from the 4 stops. This destination can subsequently be analyzed according to its location in respect to other layers. In this case a shopping area layer in which the destination is obviously inside.
All trips ending in this destination are presumed as having the purpose of shopping.

The program settings for shopping are in the example below in (see the blue dotted line in Figure 4) set as being within a distance of 100 m of a shopping area (In the figure called commercial). The setting illustrated within the red dotted line indicate that 60 % of the stops should be of the “shopping” category as defined in phase 1. The destination from Figure 3 falls within these criteria. All trips leading to this stop are therefore accepted as having the purpose of shopping.

The setting defined in the light blue column has the special feature of not splitting the trip. As described previously all time gaps above 60 seconds are considered as being a potential stop. Many of these are in the end not to be accepted as stops since they often take place because of other occurrences e.g. dense traffic. These stops are not to be split and therefore have to be treated separately.

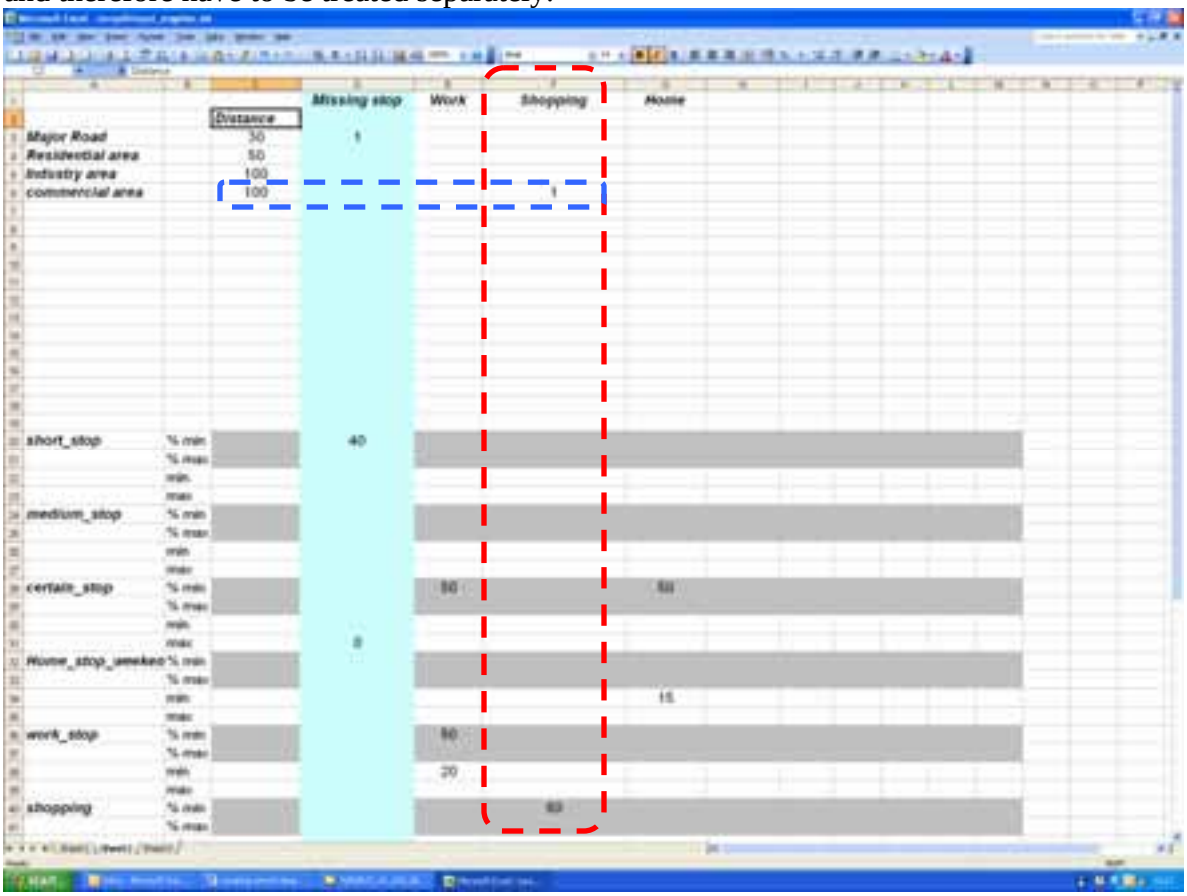


Figure 4 User interface used to define the destination

The settings used to identify these non accepeted stops are as follows.
The stop is within a distance of 30 m from a major road. This setting is based on the to the presumption that dense traffic situations occur there. There are not to be any stops of the category “certain stops” (more than 180 seconds) associated with the “destination” and at least 40 % of the registered possible stops should be of the “short stop” type.

Again with the spreadsheet as user interface, it is possible to alter these and add more definitions in any way required.

Results

The process of locating the missing stops was successful. By comparing a manual evaluation based on local knowledge with the actual result from the program, it was concluded that the program performed very well, taking the restrictions of data into account.

By comparing the program results with the real addresses given by the test persons, it was possible to measure the hit rate of this automated process. Home was matched correctly in 95 % and work in 80 % of the cases. The missed 5% and 20 % of error were generally caused by shifting workplaces in the relative short period of the experiment or simply the data regarding the addresses not being up to date.

For the other purposes such as shopping or recreational activities measuring the hit rate was more difficult. This primarily because the test persons did not inform us about the location of their other activities. In this way there were no comparable data.

Activities other than home or work are also by nature more difficult to reveal. In part because fewer stops are expected at these stops, and partially because these generally require a highly detailed land use map. But even with a very detailed map at hand, some purposes are impossible to estimate. If a gym for example is placed within or shares parking lot with a shopping area, going to the gym will probably be registered as shopping. This because it will be within the time span also expected for shopping, beside sharing the same location.

Conclusion

The available data taken into consideration, the first phase of finding the missing stops was quite successful.
In the second phase locating the home trips was the easiest and most successful part. Mainly because home is the stop where many and the longest stops are expected. At the same time the test person is in many cases expected to park in exactly the same place everyday . Work stops have some of the same characteristics and are therefore also relatively easy to distinguish. Had the persons with occupations that require extreme variance of time or place of work been handled separately from the rest, the hit rate would also in this case have been close to 100 %.
Estimating other purposes is in general much more uncertain. Applying a probability of purpose in these cases seem to be a better solution.

For similar experiments in the future the following recommendations can be given based on the experiences of this trial.
Registering the engine stop in all cases and of course avoiding technical problems would have reduced some of the uncertainties encountered.
Requiring that a sample of the test persons in some way register their trip purpose, would be a great help in the later process of adjusting the “category” and “destination” definitions. This would of course help refining the trip purpose definitions, but at the same time confirm whether a stop was intentional.
The approach that the program uses seems to be the correct one. It is only a question of refining data and definitions and the hit rate will be significantly higher.

References

[1.] Nielsen, O.ABehavioural responses to pricing schemes: Description of the Danish AKTA experiment. *Journal of intelligent Transportation Systems.* Vol 8(4). Pp.232-251. Taylor & Francis. 2004.

[2.] Nielsen, O. A.and Jørgensen, R.M. Map-matching algorithms for GPS data – Methodology and test on data from the AKTA road pricing experiment in Copenhagen. *World Conference on Transport Research Society (WCTRS),* Proceedings, Paper 1435 ,D05 July 7th, Istanbul, July 4-8, 2004