Mobile Phone Data for Telematic Applications

Prof. Dr.-Ing. Markus FRIEDRICH
Dipl.-Inf. Prokop JEHLICKA
Dipl.-Ing. Thomas OTTERSTAETTER
Dipl.-Ing., MBA Johannes SCHLAICH

Department for Transport Planning and Traffic Engineering, Universität Stuttgart
Seidenstr. 36, 70174 Stuttgart, Germany
Phone +49-711-685-84414, Fax +49-711-685-82484, Email FirstName.LastName@isv.uni-stuttgart.de

ABSTRACT

The paper reports on a research project, which applies floating phone data from mobile phones for telematic applications in transport. Floating phone data recorded in the mobile phone network are used to derive time-space trajectories of car travellers. Data recorded during the call of a mobile phone user serve as an alternative to GPS-based floating car data for a cheaper and broader traffic state recognition. Data sets from mobile phones in stand-by mode collected over a longer time period provide the basis for observing route choice behaviour and allow in combination with roadside detectors the generation of origin-destination matrices, which will be the focus of the paper. Furthermore, the data may be used for a continuous monitoring of the service quality and the travel demand in road networks.

Keywords:
Floating phone data, traffic state recognition, route choice behaviour, origin-destination matrices, origin-destination matrix projection, data-clustering

1 INTRODUCTION

The implementation of state-of-the-art traffic control systems requires methods for detecting the current traffic state. A high quality traffic state recognition is a prerequisite for providing reliable traveller information, for forecasting the future state of the transport network and, in combination with a model, for evaluating and implementing control strategies. The primary data sources for traffic state recognition are roadside detectors (e.g. loop detector, overhead detector) which can measure occupancy rates, volumes and speeds for one lane of a road. Mobile detectors moving along with the traffic flow complement the stationary data by delivering time-space trajectories of single travellers. Usually mobile detectors are integrated in the navigation system of cars which broadcast floating car data (FCD) to a control centre. Main disadvantages of FCD are the low rate of equipped vehicles and the still relatively high communication costs. As an alternative, observation of mobile phones can provide floating phone data (FPD or Net-FCD) for a large number of travellers in the network. Implementations by TomTom Mobility Solutions [1], Vodafone [2] and others show that FPD are a considerable data source for traffic state detection.

The research project Do-iT, funded by the German Federal Ministry of Economics and Technology, aims at exploiting FPD in a pilot area in Southern Germany. The research project which is scheduled to run until 2009 intends to prove that FPD contribute to the following objectives:

- Provision of traffic state information for motorways and the secondary road network.
- Improving methods for detection and forecast of the traffic state by applying origin-destination matrices derived from FPD.
- Continuously monitoring the service quality (travel times) and the travel behaviour (travel demand, route choice) in the network for planning purposes and for evaluating telematic control strategies.

2 MOBILE PHONE BASED DATA COLLECTION

The collection of the relevant mobile phone information uses only existing communication protocols and data flow within the mobile phone network. As shown in Figure 1 this requires the installation of network probes to record the relevant data at appropriate interfaces of the mobile phone network infrastructure. During the call of a mobile phone user detailed data on the reception intensity of the currently used radio cell and of the neighbouring cells are recorded at the A-bis-level. In contrast the A-level interface records only the hand-over from one cell to the next cell during a conversation or the change of the location area, which consists of roughly 20 cells, if the mobile phone is in the stand-by mode.
Processing the FPD from the network probes requires two steps:

- The identification of mobile phone users who represent car travellers. This requires algorithms, which distinguish between mobile phones located in cars and in trains, even if the railway runs parallel to the motorway.
- The generation of time-space trajectories using modified map-matching algorithms.

On the A-bis level these two steps come down to the following procedures. First, the mobile phone network positions are transformed to coordinates in the street map based localization network, using signal strengths and signal strength maps. The resulting positions are then smoothed with the help of a Kalman filter [3]. After the identification of mobile phones that are actually moving, a map matching algorithm is applied. In order to identify the moving mobile phone’s means of transport, its trajectory is being compared with the public transport timetables by a fuzzy based algorithm. Depending on the fact whether the resulting degree of truth surpasses a certain threshold, the mobile phone is identified as a public transport participant. The individual procedures of the FPD generating process are described thoroughly by Wiltschko and Schwieger [4]. More literature related to the generation and identification of FPD can be found in [5], [6] and [7].

After a trajectory for a single phone call has been generated, the algorithm tries to connect other possibly available trajectories of the same mobile phone. At this point the processing of the less detailed raw data on the A level starts. By identifying and connecting definitely continuous trajectories (e.g. the first one ends on a motorway and the next one starts 5 minutes later and 9 kilometres further on the same motorway) or cell changes, it is possible to extract FPD that are valid for a longer period and distance, representing valuable information for traffic applications.

### 3 APPLICATION OF FLOATING PHONE DATA

Figure 2 shows three major traffic applications for which routes and trajectories from FPD may be used as input:

- recognition of the current traffic state,
- route choice analysis and
- generation of origin-destination (OD) matrices.

Until now the recognition of the current traffic state is the main objective of traffic service providers and network operators. For corridor control systems, for the monitoring of the service quality and for planning purposes observing the route choice and journeys from their origin to their destination is a further objective. This paper will focus on the second objective.

### Observing Trajectories of Journeys

To detect the traffic state it is adequate to examine the travel times along relatively short sections of a journey. In order to deliver information about the route choice of a traveller or the origins and destinations of journeys it is necessary to identify the complete journey. As barely any traveller will talk on his mobile phone for the entire duration of his journey the Location Area Updates in the standby-mode are of special importance. At any change of a Location Area the current radio cell is recorded. The spatial extension of Location Areas in the test field varies in dependence of the location and the direction of transition between 5 km in urban areas and 25 km in rural areas. Thus when covering the 65 km section from the motorway interchange Stuttgart to the motorway junction Karlsruhe six Location Areas are passed. Considering these constraints the projects aims at determining journeys with a travel length of more than 20 km.
**Route Choice**

For the first time FPD can provide continuous observations of travel behaviour with respect to route choice. Observations of revealed route choice decisions can be combined with measured indicators relevant to individual route choice, e.g. current travel time, historical travel time, state of variable message signs or TMC-RDS messages. The comprehensive data set from FPD will be analysed with maximum-likelihood-estimates. This allows the calibration of route choice models which may serve for planning purposes as well as for operating control systems in the motorway network.

Figure 3 shows exemplarily the route choice on two alternative routes in a part of the German motorway network. The presented shares are recorded with automated number plate recognition systems and serve as a source for validating the FPD. Both routes connect the motorway interchanges Stuttgart and Walldorf. Route 1 via Karlsruhe is 95 km long, route 2 via Heilbronn is slightly longer. Figure 3 also shows the traffic state reported by the traffic message channel. It can be seen that the car drivers react to the broadcasted traffic news, which report heavy congestions on the route via Karlsruhe. No disturbances are reported for the second route.

**Generation of Origin-Destination Matrices**

The continuous recording of travel behaviour allows generating OD matrices and travel time matrices with a high temporal resolution. The OD matrices of mobile phone users can be projected to demand matrices of all car travellers using traffic counts from roadside detectors. This is necessary because the amount of FPD does not correspond directly to the larger number of actual vehicle trips - it may not have been possible to generate trajectories for every moving mobile phone, not every vehicle carries a mobile phone and not every mobile subscriber is under contract with the mobile company which delivers the data.

Existing matrix correction methods are unable to use time dependant traffic counts as a performance function and have to rely on traffic assignment methods to evaluate their results. In order to take full advantage of FPD, regarding the generation of OD matrices, a dynamic projection method has been developed.

In order to generate OD matrices for typical days (working day, Saturday, Sunday, start of holiday) two approaches for clustering the demand matrices are examined (Figure 4). Both approaches are based on traffic counts of the stationary detectors and on the mobile phone trajectories that are available for a longer period of calendar days. The trajectories of the mobile phones start and end in a traffic zone which corresponds to a Location Area.

1. In the first approach the mobile phone trajectories are projected for every single calendar day. For this projection the roadside detector counts of the respective calendar day are used. As a result for every single mobile phone trajectory a specific projection factor is obtained that shows how many vehicles are represented by this trajectory. The OD matrix of a calendar day can now be derived directly from the origins and the destinations of the projected trajectories. In order to determine typical OD matrices from the large number matrices by calendar day the matrices are clustered.

2. In the second approach the roadside detector counts are clustered in order to obtain typical traffic days. A traffic day includes several calendar days. Now the mobile phone trajectories of all calendar days covered by a traffic day
are projected in such a way that the traffic volumes resulting from the trajectories comply with the mean traffic volumes of the stationary detectors.

The second method seems to be more suitable as it reduces the number of time-intensive projections. In addition, the accumulated FPD-trajectories over all days of the respective cluster represent a robust input for the projection method.

The applied cluster methods are based on the single/average cluster represent a robust input for the projection method. The second method seems to be more suitable as it reduces the accumulated FPD-trajectories over all days of the respective number of time-intensive projections. In addition, the position dependant function, which has been used by Vortisch [9] and the GEH-function, named and invented by Geoffrey E. Havers [10], which has been specifically designed to express meaningful differences in traffic flows.

Theory

While time-dependant FPD-OD matrices can be generated by simply summing up all FPD-trajectories within a certain time interval, the generation of demand matrices corresponding to actual trips is a more complex task. The unknown and time- or area-dependant average number of mobile phones per vehicle implicates the need for a specific projection method. Figure 5 shows the schematical layout of this process.

The dynamic projection is an iterative process, with the following paragraphs describing the steps of a single iteration.

The procedure uses all available FPD-trajectories and roadside detectors as input and calculates a correction factor $C_i$ for the number of trips every single trajectory $i$ reflects. This correction factor is based on the average ratio of all roadside counted traffic flows $Q_{\text{count}}(t,p)$ alongside the FPD-trajectory and the sum of all FPD-trajectories $Q_{\text{FPD}}(t,p)$ at the corresponding discrete time and position intervals $t$ and $p$:

$$C_i = \frac{1}{n} \sum_{j=1}^{n} \frac{Q_{\text{count}}(t,p)}{Q_{\text{FPD}}(t,p)}$$  

with

- $C_i$ correction factor for trajectory $i$
- $n$ number of detectors alongside trajectory $i$
- $j$ roadside detector $j$
- $Q_{\text{count}}(t,p)$ counted traffic flow for the time interval $t$ at the position $p$
- $Q_{\text{FPD}}(t,p)$ summed up FPD traffic flow for the time interval $t$ at the position $p$

Trajectories that do not run over any roadside detectors are projected with the average correction factor $C_{\text{avg}}$ that corresponds to the total trip change of all FPD-trajectories that do run over roadside detectors:

$$C_{\text{avg}} = \frac{Q_{\text{new total}}}{Q_{\text{old total}}}$$

The iterative projection process needs to be able to handle a various number of possible contradictory boundary conditions, e.g.:

- Roadside detectors that discontinue working properly. Two detectors alongside an enclosed section of a motorway - implying no possibility for vehicles to appear or disappear - that are reporting different vehicle counts demand an impossible solution. In this case the algorithm tends to converge slowly to a solution between these counts. The exact value depends on the counts and locations of other detectors that are nearby, but is very likely not able to satisfy a target maximum error.

- Areas of the network that are not covered by FPD at all times. Streets counted by roadside detectors, that are not being passed by any FPD-trajectory demand also an impossible solution. Because the number of trips a FPD-trajectory reflects is multiplied with a correction factor, it is unfeasible to generate trips out of nothing. One has to keep in mind that this dilemma does not necessarily correspond to regions with no signal reception. The pilot area for the project Do-iT is well covered by the mobile phone network, but not every vehicle is carrying a mobile phone that is using a T-Mobile contract (the raw data supplier). Additionally it is not possible to generate FPD-trajectories for short distance trips and due to a huge demand in computing power a lot of data remains unused.

The iteration process is controlled by the following variables:

- Maximum number of allowed iterations.
- Reaching a target maximum error.
- Undershooting minimal changes for significant parameters for subsequent iterations.
Analysis

Due to a lack of real data at the current state of the project the dynamic projection has been subject to several simulation based tests. One of these will be discussed in the following paragraphs.

Figure 6 shows the rather simple but comprehensible test network. The network consists of six links, seven nodes, five zones and the necessary connectors. Traffic demand is flowing from left to right, fluctuating over the simulation time of 60 minutes and reaching its peak between the minutes 35 and 50. The solid black lines depicted in Figure 7 display the load curves on the links 2, 4 and 6 that are the result of a microscopic traffic simulation. The plots are discretized in steps of one minute, because this will be the time interval the dynamic projection looks at. The microscopic simulation is able to generate detailed trajectories for every single vehicle and will represent the reality for the analysis as well as complement the dynamic projection with “FPD”-trajectories.

In order to reflect the fact, that FPD-trajectories generated out of mobile phone data will only represent a slice of the trajectories of every single vehicle, the input “FPD”-trajectories for the dynamic projection will be reduced to 359 trajectories. These represent about 12.5% of the total 2870 trajectories. The resulting FPD traffic flow is depicted by the dashed grey lines in Figure 7.

The input for the dynamic projection will be these 359 “FPD”-trajectories and the load curves on links 2, 4 and 6 that will represent counted data generated by roadside detectors. After ten iterations the maximum change in the number of vehicle trips each FPD-trajectory represents drops below 1%, which terminates the projection process. The resulting FPD-flows for the considered links 2, 4 and 6 are plotted in Figure 8, again in combination with the counted respectively simulated data.

Barring one deviation on link 4 during minute 59 and two deviations on link 6 during the minutes 40 and 57 the curves are almost congruent and therefore the result of the dynamic projection can be considered as very satisfying.

Further analysis of the input data and the process itself explains the reasons for the aforementioned deviations:

- Figure 8, Link 6 (bottom plot) minute 57:
  As shown by Figure 7 the starting conditions are unsuitable to obtain the counted flows. The FPD-flows equal zero from the beginning. As the projection tries to change the number of trips by multiplying the starting conditions, it is...
impossible to reach the target value. This corresponds to the aforementioned unfortunate case of areas that are not being covered by FPD.

- Figure 8, Link 6 (bottom plot) minute 40 and link 4 (middle plot) minute 59:
  These two deviations are interconnected. During minute 40 there are exactly two trajectories passing through link 6. The same two trajectories reach the downstream link 4 during minute 59 together with three other trajectories. Without the possibility of a negative number of vehicle trips per trajectory it becomes easily understandable, that it is not possible to fulfill a high traffic count at link 6 minute 40 and a low traffic count at link 4 minute 59 at the same time. The reason for that contradictory requirement is similar to the previous one. By chance the FPD-trajectories that would have enabled the satisfaction of these boundary conditions were not generated and are therefore not available as input for the dynamic projection.

Despite of these already identified possible flaws the result of the projection process can be valued as high-quality. The influence and dimension of necessary FPD-trajectories, that are missing as input, on the practical application of the dynamic projection can only be determined after a large amount of real data has been processed.

4 CONCLUSION

The development of FPD as a cost-effective source for traffic data paves the way for the extension respectively the replacement of other survey methods and the continuous monitoring of the actual traffic behaviour opens up new ways in research, planning and operation of transport.

The dynamic projection method that has been developed is a promising tool to generate respectively to correct OD matrices. The method does not rely on traffic assignments, therefore reducing the influence of other models. It also takes into account all available time dependant date, hence giving the possibility to respect the influence of traffic jams on demand data.

5 REFERENCES


