Available online at www.sciencedirect.com
ScienceDirect
Procedia Computer Science 37 (2014) 390 – 395
Procedia Computer Science

First International Workshop on Information Fusion for Smart Mobility Solutions (IFSMS14)

Floating Car and Camera Data Fusion for Non-Parametric Route Travel Time Estimation

Mahmood Rahmani[a,*], Erik Jenelius[a], Harilaos N. Koutsopoulos[a]

[a] *Department of Transport Science, KTH Royal Institute of Technology. Teknikringen 72, SE-100 44 Stockholm, SWEDEN*

Abstract

Traffic management centers take advantage of various data collection systems ranging from stationary sensors e.g. automated vehicle identification systems to mobile sensors e.g. fleet management systems. Each type of data collection system has its own advantages and disadvantages. Stationary sensors has less measurement noise than mobile sensors but their network coverage is limited. On the other hand, mobile sensors cover expand areas of road networks but they have less penetration rate and frequency of reports. Traffic state estimation can benefit from fusion of data from various sources as they complement each other. This paper introduces a route travel time estimation method that aggregates data from two traffic data sources, automated number plate recognition system and floating car data.

Peer-review under responsibility of the Program Chairs of EUSPN-2014 and ICTH 2014.

Keywords: route travel time estimation; traffic data fusion; floating car data; automatic vehicle identification;

1. Introduction

In light of increasing congestion in urban areas, monitoring and providing information about traffic conditions is critical for traffic management and effective transport policy. Travel time data may be collected from stationary automatic vehicle identification (AVI) sensors (automatic number plate recognition (ANPR) cameras, Bluetooth devices, etc.). AVI systems provide direct measurements of route travel times, but the spatial coverage is typically small and may not be representative of the network as a whole. Meanwhile, floating car data (FCD) collected from GPS devices installed in vehicle fleets or smart phones provide information from the entire network. Travel time estimation from FCD is often challenging because of low penetration rate, which means that the number of available FCD observations from vehicles traveling along the route of interest may be low if it is not a common one.

AVI and FCD have complementary strengths as FCD provides network coverage while AVI provides accurate measurements on specific route segments. The combination of AVI data and FCD has not been studied much in the literature, however. A data fusion methodology for freeway traffic state estimation based on loop detector data, AVI and FCD has been proposed[1]. For the arterial network, research on travel time estimation based on FCD has largely focused on links[2,3].

* Corresponding author. Tel.: +46-8-790-8301 ; fax: +46-8-21-2899.
E-mail address: mahmoodr@kth.se

1877-0509
Peer-review under responsibility of the Program Chairs of EUSPN-2014 and ICTH 2014.
doi:10.1016/j.procs.2014.08.058

The aim of this paper is to utilize the complementary benefits of ANPR and FCD by integrating the two data sources in the estimation of arterial route travel times. The paper proposes a computationally efficient, non-parametric method for route travel time estimation using both ANPR data and low-frequency FCD. The approach estimates route travel time distributions directly from ANPR and FCD measurements partially covering the route, incorporating all information available from the data. The methodology extends ideas from kernel-based estimation[4] and is developed considering the particular features of network routes and ANPR and FCD observations. No assumptions are made regarding the form of the distribution. This flexibility is highly valuable whenever the variability of travel times is of interest, e.g., for monitoring of travel time reliability.

The paper is organized as follows: Section 2 describes the methodology, Section 3 presents a case study for Stockholm, Sweden, and Section 4 concludes the paper.

2. Methodology

2.1. Preliminaries

A *network route* is defined as an acyclic path $\pi = (k_s, k', \ldots, k'', k_e)$ connecting the beginning and end links k_s and k_e, and two distances o_s and o_e marking the two offsets on k_s and k_e respectively. The fraction of link k covered by the network route, denoted α_k, is the length of overlap between the route and the link divided by the link length. The route travel time is denoted by $T = T(s)$ and varies stochastically between trips and as a function of the route entry time s. The aim of this research is to estimate the distribution of $T(s)$ from ANPR and FCD measurements.

Automatic number plate recognition (ANPR) data. ANPR data are collected from ordered pairs of cameras which identify vehicles based on optical recognition of license numbers. An ANPR route is defined as the path between the locations of the first and the second camera (more precisely, the locations where vehicles are detected); it is assumed that there is a single reasonable route between the two locations. A data record is created whenever the same vehicle is identified sequentially by both cameras. A record is a triplet (h, s, e), where h is a unique ANPR route identifier, and s and e are the timestamps of the detection of the vehicle at the first and the second camera, respectively.

Floating car data (FCD). FCD consist of sequences of reports, or *probes*, from vehicles traveling on the network. Each probe is a triplet $(q, s, < x, y >)$, where q is a unique vehicle identifier, s is a timestamp and $< x, y >$ are the GPS coordinates of the vehicle location at that time. Low-frequency FCD require preprocessing to be useful for travel time estimation. Most importantly, reported positions must be matched to the model of the road network and the paths taken by the vehicles between probes must be inferred[5,6].

2.2. Travel time estimation

A computationally efficient non-parametric method for route travel time estimation from FCD has recently been developed[7]. This paper extends the method to combine FCD with available ANPR data overlapping the route. A common observation model, consisting of a route and a travel time measurement, is used to represent both ANPR data and FCD. For ANPR observations the route is given by the fixed camera locations and the intermediate route, and the travel time measurement $\tau = e - s$ is obtained from the difference between the detection timestamps. For FCD observations the route is given by the inferred path between two consecutive probes from the same vehicle, and the travel time measurement $\tau = s_2 - s_1$ is obtained from the difference between the corresponding timestamps. The framework is illustrated in Figure 1.

In general, observation i from either ANPR or FCD is represented by a travel time τ_i, a path $p_i = (k_{i,1}, k', \ldots, k'', k_{i,2})$ and two distances $o_{i,1}$ and $o_{i,2}$ marking the two offsets on $k_{i,1}$ and $k_{i,2}$ respectively. The fraction of link k traversed by measurement i is denoted by ρ_{ik}. The part of the observation route overlapping with the network route is referred to as the *overlap route* for short. The fraction of the overlap in relation to the length of the link is denoted by β_{ik}.

ANPR and FCD observations are processed together in three steps: transformation, weighting, and aggregation.

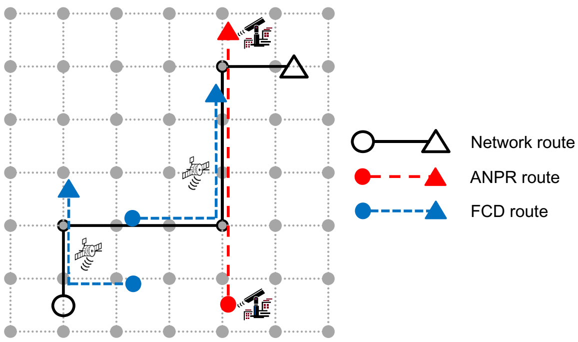

Figure 1. Observations from ANPR and FCD overlapping the network route.

Transformation. Each observation partially covering the route is transformed into an observation of the actual route travel time. The step consists of four sub-parts: concatenation, allocation, scaling, and route entry time estimation. Concatenation applies to FCD and sequences of ANPR cameras, where a vehicle may generate multiple data records along the route. It is reasonable, however, to consider one passage of a vehicle on the route as one travel time observation. Consecutive observations from the same vehicle are thus concatenated into a single travel time observation.

For each observation i, the observed travel time τ_i is then allocated between the network route and the adjacent network. The allocation is based on prior link travel times t_k^0 and the distance traversed on each link. The assumption is that the fraction of time spent on the overlap route in relation to the whole FCD route, ϕ_i, is the same as for the prior travel times on the same sections. The travel time allocated to the network route is then $\tau'_i = \phi_i \tau_i$, where the allocation factor ϕ_i is $\phi_i = \sum_k \beta_{ik} t_k^0 / \sum_k \rho_{ik} t_k^0$.

The travel time observations are then scaled up to the entire route. Similar to the allocation, the assumption is that the ratio between the travel time on the overlap route and on the entire network route is the same as for the prior travel time estimates on the same sections. The scaled route travel time observation is then $T_i = \tau'_i / \eta_i$, where η_i is the scaling factor $\eta_i = \sum_k \beta_{ik} t_k^0 / \sum_k \alpha_k t_k^0$.

The time that each observed vehicle passes the beginning of the network route is the basis for grouping observations to time intervals and aggregating, but is in general not observed. For each observation the route entry time s', real or hypothetical, is estimated based on the prior travel times along the same lines as the allocation and the scaling.

Weighting. Each observation T_i is assigned a weight ω_i that determines its influence in the estimation of route travel time statistics. Observations are weighted for three distinct reasons: to reflect the representativeness in relation to the network route; to correct for sampling bias due to uneven route coverage; and to reflect the relative reliability of FCD and ANPR measurements. The final weight is the product of the representativeness weight ν_i, the sampling bias weight λ_i and the source reliability weight γ_i, i.e., $\omega_i = \nu_i \lambda_i \gamma_i$. In this paper γ_i is set to 1 for both ANPR and FCD observations.

Less overlap with the network route means that the representativeness as a network route observation is lower and that the potential for error in allocation and scaling is higher. Observations are thus weighted based on the allocation and scaling factors, $\nu_i = \phi_i^{1/\theta_1} \eta_i^{1/\theta_2}$. The parameters θ_1 and θ_2 control how fast the weight function decays as the overlap with the adjacent networks increases, and the overlap with the network route decreases, respectively. Both parameters are set to 1 in this paper. If ground-truth travel time data were available, they can be selected using cross-validation techniques.

Route coverage is evaluated at the link level. Let N_k be the number of observations covering link k. The weight λ_i is then the inverse of the weighted average coverage for the traversed part of the route, $\lambda_i = \sum_k \beta_{ik} t_k^0 / \sum_k \beta_{ik} t_k^0 N_k$.

Aggregation. Statistics of the route travel time distribution are calculated from the observations T_i and the associated weights ω_i. The statistics are aggregated based on the route entry time of each observation. For example, the mean route travel time estimator is $\hat{\mu}_T = \sum_i \omega_i T_i / \sum_i \omega_i$. Other statistics of the travel time distribution such as variance and percentiles are also straightforward to calculate[7].

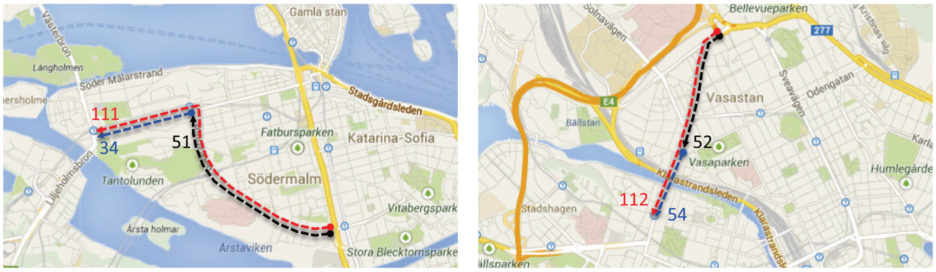

Figure 2. The study area with 4 ANPR routes (51, 34, 52, and 54), and 2 combined routes, 111 and 112.

3. Application

The proposed travel time estimation method is applied on two routes in the arterial network of Stockholm, Sweden. FCD are collected by a GPS-based taxi fleet management system covering about 1500 taxis. Each taxi broadcasts its location, timestamp, id number, and status (free/hired) once every two minutes on average[8]. A map-matching and path inference algorithm is performed on the raw FCD[6]. The ANPR system in Stockholm measures direct travel time of many major routes. ANPR data typically have significant amounts of noise and need to be filtered before travel time statistics are calculated. The method introduced by Kazagli et al.[9] is used for ANPR filtering.

For both FCD and ANPR, data from Mondays through Thursdays between 6 a.m. and 10 p.m. are used, collected from September 15, 2012 to September 15, 2013. Only FCD with *hired* status is utilized for travel time estimation. The prior link travel times t_k^0 are estimated from FCD by applying the proposed method on each individual link. The allocation, scaling and weighting of FCD in this step are performed based on the length and free-flow speed (posted speed limit) of each link.

3.1. Experimental setup

Two network routes are defined for the experiment, denoted as routes 111 and 112 in Figure 2. Route 111 (112) starts from the beginning of ANPR route 51(52) and ends at the end of ANPR route 34(54). The network routes are intentionally defined this way to be the combination of the two consecutive ANPR routes for comparison purposes. Since consecutive ANPR routes r_1 and r_2 share the same camera at their connection point, direct travel time observations for the combined route R are available by matching the exit timestamp of r_1 with the entry timestamp of r_2. The direct ANPR observations are used as reference for evaluating the travel time estimation.

The mean and the 25th, 50th (median) and 75th percentiles of the travel time distribution of route R is estimated using FCD and ANPR data considering the following scenarios:

- *Only FCD* that fully or partially overlap the route R
- *Only ANPR of* r_1
- *Only ANPR of* r_2
- *ANPR of* r_1 *and FCD*
- *ANPR of* r_2 *and FCD*

Travel time observations are grouped by route entry timestamp into 15-minute intervals, and the travel time statistics are calculated for each time interval ($N = 64$ intervals from 6 a.m. to 10 p.m.). The similarity between the estimated travel time statistic and the reference based on direct observations is evaluated using the root-mean-square error (RMSE) across all time intervals.

3.2. Results

Figure 3 shows the estimated travel time of the two routes, 111 and 112, under the data availability scenarios mentioned above. In general, the estimation method capture the trend across the day reasonably well in all scenarios, in particular for the mean and median travel time. The RMSE relative to the direct measurements, as well as the number of observations available for the estimation, are shown in Tables 1 and 2. It can be seen that "only FCD" performs better than "only ANPR" in some cases but worse in others, and the results of "only ANPR" differ significantly depending on which ANPR route is used. The estimator fusing ANPR and FCD is always better than the worst of "only ANPR" and " only FCD", and in some cases is better than the best of the two. The results suggest that the fusion of ANPR and FCD increases the robustness of the estimation.

Table 1. Error of estimated travel time of route 111 by various scenarios against the "Only ANPR" scenario.

	Only FCD_{111}	Only $ANPR_{51}$	Only $ANPR_{34}$	$ANPR_{51}$ and FCD_{111}	$ANPR_{34}$ and FCD_{111}
mean	0.28	0.46	0.26	0.34	0.23
25^{th} percentile	0.84	0.72	0.73	0.74	0.77
50^{th} percentile	0.49	0.50	0.33	0.46	0.34
75^{th} percentile	0.26	0.29	0.53	0.22	0.45
Total number of observations	84,590	103,423	1,001,696	188,013	1,086,286

Table 2. Error of estimated travel time of route 112 by various scenarios against the "Only ANPR" scenario.

	Only FCD_{112}	Only $ANPR_{52}$	Only $ANPR_{54}$	$ANPR_{52}$ and FCD_{112}	$ANPR_{54}$ and FCD_{112}
mean	0.30	0.16	0.53	0.16	0.44
25^{th} percentile	0.33	0.22	0.68	0.23	0.60
50^{th} percentile	0.18	0.16	0.28	0.14	0.19
75^{th} percentile	0.37	0.25	0.89	0.25	0.71
Total number of observations	95,944	305,227	768,577	401,171	864,521

4. Conclusion

The paper proposes a non-parametric route travel time estimation method based on fusion of FCD and ANPR data. The approach combines the network coverage of FCD with the accurate measurements on specific route segments of ANPR. A common observation model for both sources of data is used to estimate travel time through a sequence of transformation, weighting and aggregation. Application results suggest that the fusion increases the robustness of the estimation, meaning that the fused estimate is always better than the worst of the two (FCD or ANPR), and sometimes better than the best of them. Further research is needed to evaluate the method on a varied set of network routes and data sources, and to calibrate the parameters of the method to optimize the performance.

References

1. J. W. C. van Lint, S. P. Hoogendoorn, A robust and efficient method for fusing heterogeneous data from traffic sensors on freeways, Computer-Aided Civil and Infrastructure Engineering 25 (2010) 596–612.
2. A. Hofleitner, R. Herring, P. Abbeel, A. Bayen, Learning the dynamics of arterial traffic from probe data using a dynamic Bayesian network, IEEE Transactions on Intelligent Transportation Systems 13 (2012) 1679–1693.
3. F. Zheng, H. van Zuylen, Urban link travel time estimation based on sparse probe data, Transportation Research Part C 31 (2012) 145–157.
4. T. Hastie, R. Tibshirani, J. Friedman, The Elements of Statistical Learning, 2nd Edition, Springer, 2009.
5. T. Miwa, D. Kiuchi, T. Yamamoto, T. Morikawa, Development of map matching algorithm for low frequency probe data, Transportation Research Part C 22 (2012) 132–145.
6. M. Rahmani, H. N. Koutsopoulos, Path inference of sparse GPS probes for urban networks, Transportation Research Part C 30 (2013) 41–54.
7. M. Rahmani, E. Jenelius, H. N. Koutsopoulos, Non-parametric estimation of route travel time distributions based on low-frequency floating car data, Tech. rep., KTH Royal Institute of Technology, Dept. of Transport Science (2014).
8. M. Rahmani, H. Koutsopoulos, A. Ranganathan, Requirements and potential of GPS-based floating car data for traffic management: Stockholm case study, in: Intelligent Transportation Systems (ITSC), 2010 13th International IEEE Conference on, IEEE, 2010, pp. 730–735.
9. E. Kazagli, H. N. Koutsopoulos, Arterial travel time estimation from automatic number plate recognition data, Proceedings of the 92nd Annual Transportation Research Board Meeting (2013).

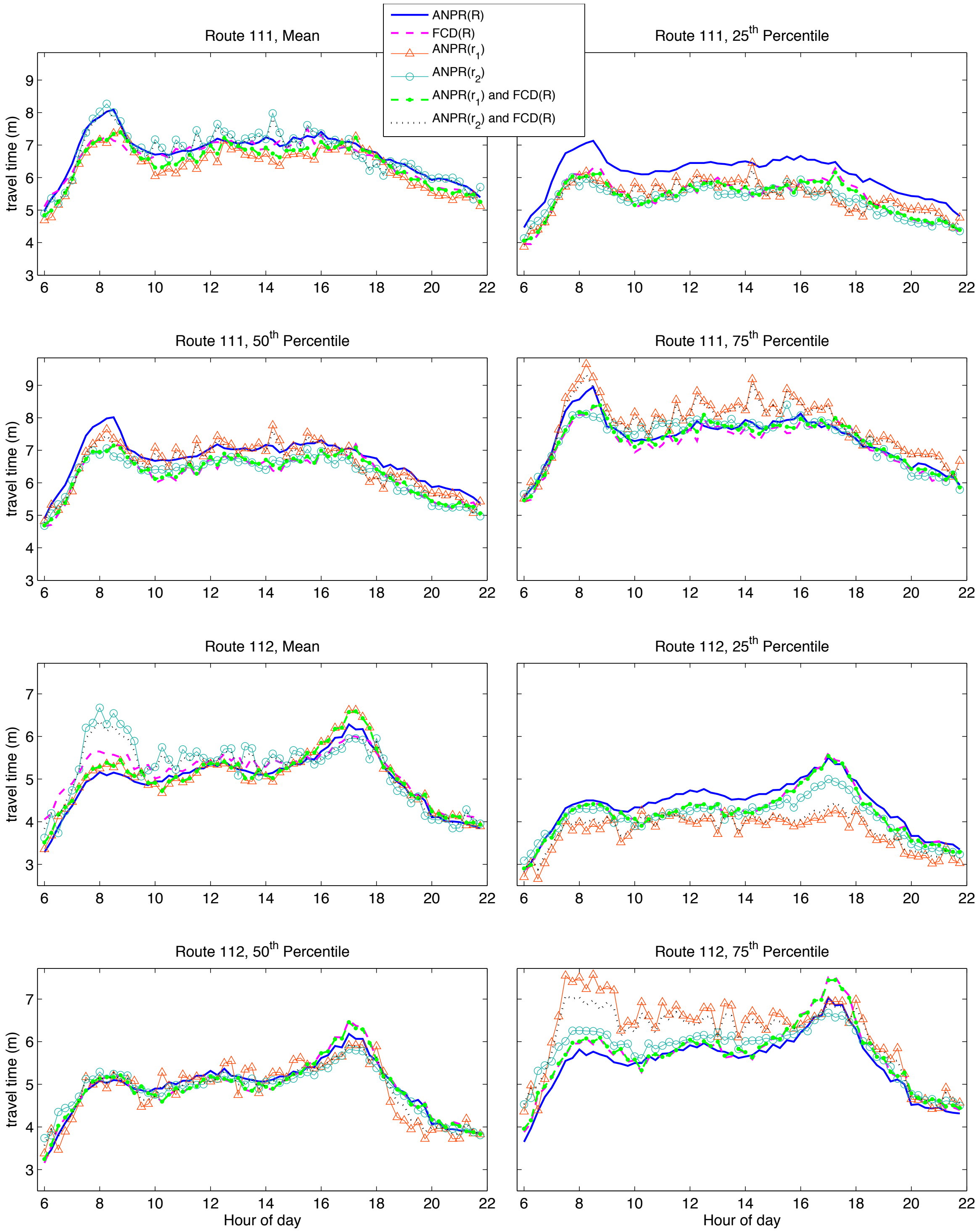

Figure 3. Comparison of estimated mean and percentiles of the travel time of routes 111 and 112 by various scenarios against direct ANPR observations.